

# A Gene Regulatory Model of Cortical Neurogenesis

Sabina S. Pfister<sup>a</sup>, Andreas Hauri<sup>a</sup>, Frederic Zubler<sup>a,b</sup>, Gabriela Michel<sup>a</sup>, Henry Kennedy<sup>c</sup>, Colette Dehay<sup>c</sup>, Rodney J. Douglas<sup>a,\*</sup>

<sup>a</sup>*Institute of Neuroinformatics, University of Zurich and ETH Zurich, 8057, Zurich, Switzerland*

<sup>b</sup>*Department of Neurology, Bern University Hospital, 3010, Bern, Switzerland*

<sup>c</sup>*University of Lyon, Université Claude Bernard Lyon 1, Inserm, Stem Cell and Brain Research Institute U1208, 69500 Bron, France.*

---

## Abstract

Sparse data describing mouse cortical neurogenesis were used to derive a model gene regulatory network (GRN) that is then able to control the quantitative cellular dynamics of the observed neurogenesis. Derivation of the network begins by estimating from the biological data a set of cell states and transition probabilities necessary to explain neurogenesis. We show that the stochastic transition between states can be implemented by the dynamics of a GRN comprising only 36 abstract genes. Finally, we demonstrate using detailed physical simulations of cell mitosis, and differentiation that this GRN is able to steer a population of neuroepithelial precursors through mitotic expansion and differentiation to form the quantitatively correct complex multicellular architectures of mouse cortical areas 3 and 6. We find that the same GRN is able to generate both areas though modulation of only one gene, suggesting that arealization of the cortical sheet may require only simple improvisations on a fundamental gene network. We conclude that even sparse phenotypic and cell lineage data can be used to infer fundamental properties of

---

\*Corresponding Author

Email address: [rjd@ini.uzh.ch](mailto:rjd@ini.uzh.ch) (Rodney J. Douglas)

neurogenesis and its organization.

*Keywords:* development, neocortex, cortical cell lineage

---



## 1. Highlights

- Estimation of the cell states and transition probabilities of neurogenesis from experimental data.
- Design of an abstract gene regulatory network (GRN) whose dynamics implement cell states and their stochastic transitions.
- Detailed simulation of GRN-guided neurogenesis for mouse cortical areas 3 and 6.
- Different dynamics of neurogenesis of distinct cortical areas arise through modulation of only a single gene.

## 2. In brief

Pfister et al. show how sparse phenotypic and cell lineage data can be used to infer a small abstract gene regulatory network (GRN), which, when inserted into model precursor cells, is able to control in a distributed manner the quantitative cellular dynamics of neocortical neurogenesis.

### 15 **3. Introduction**

16 Unlike human engineered systems that are explicitly designed and constructed,  
 17 the rules for self-construction of biological organisms are implicit in the information  
 18 contained in their initial cells. Although many details of this remarkable process  
 19 have been described experimentally, there are as yet no detailed generative models  
 20 that describe formally the principles of control and global coherence amongst  
 21 proliferating, locally independent, cellular agents. Here we describe a number  
 22 of significant advances toward this goal in the context of the development of the  
 23 laminated neocortex from its neuroepithelial precursors. We show how sparse  
 24 phenotypic and cell lineage data can be used to infer a small abstract gene network,  
 25 which, when inserted into model precursor cells, is able to steer in a distributed  
 26 manner the quantitative cellular dynamics of neocortical neurogenesis. Our results  
 27 offer an insight into principles of physical self-construction of biological neural  
 28 networks.

29 Neocortical pyramidal cells are generated, and migrate to form a type specific  
 30 lamination, however, the cellular mechanisms that underly this cortical neurogene-  
 31 sis remain elusive (Greig et al., 2013). Cortical neurogenesis begins from a sheet  
 32 of neuroepithelial stem cells. These cells differentiate predominantly into radial  
 33 glial cells (RGC) (Hartfuss et al., 2001; Miyata et al., 2001; Noctor et al., 2001,  
 34 2002; Anthony et al., 2004). RGCs divide at the apical surface of the ventricular  
 35 zone (VZ), where they undergo stereotypical sequences of cell divisions: Sym-  
 36 metric divisions lead to similar offspring and amplify the pools of precursor cells;  
 37 asymmetric divisions give rise either to various intermediate precursors, (Franco  
 38 and Müller, 2013; Guo et al., 2013), or directly to cortical neurons (Heins et al.,  
 39 2002; Malatesta et al., 2003; Anthony et al., 2004; Cárdenas et al., 2018) (reviewed

in Götz and Huttner (2005)). Some precursors are restricted to the VZ (Haubensak et al., 2004; Miyata et al., 2004; Noctor et al., 2004), and are the major source of the deep layer pyramidal neurons. Other precursors form a second germinal layer, the subventricular zone (SVZ). There they undergo a few rounds of symmetric division and generate neurons largely fated for the superficial layers (Noctor et al., 2004; Kowalczyk et al., 2009).

The genealogical lineages whereby the neuroepithelial stem cells give rise to differentiated neurons are only partially known (Haydar et al., 2003; Noctor et al., 2004; Gao et al., 2014; Vasistha et al., 2015; Telley et al., 2016; Beattie and Hippenmeyer, 2017; Kaplan et al., 2017; Zhong et al., 2018). Every cell in the lineage has the same genotype, but the phenotype of each cell is due to its particular gene expression pattern, and interaction with environmental factors. The lineage tree describes the genealogy and division history of successive precursors, where each cell is associated with a particular phenotype. Ideally, the structure of the lineage tree should reflect the progressive restriction of cell fate. It would exhibit the variety of successive precursors that could be generated as neurogenesis proceeds, and thereby offers insights into the mechanisms that lead to the generation of experimentally observed neural cell types.

Although recent work points to an orderly and deterministic proliferation, and neurogenic behavior of precursors (Gao et al., 2014), the underlying organization of their lineage trees are not completely known. In principle, the progression of cell types through the tree can be characterized by their phenotypic description. The overall phenotype of a given cell can be represented as a vector of features  $f = \{f_1, f_2, \dots, f_n\}$  that include its gene expression pattern, morphology, biochemical or physiological properties, and behavior. Some of these features may be

65 observable, but others are hidden. We assume that this vector of cell features  
66 is conditioned by the internal unobservable cell state  $S$  that completely explains  
67 their distribution. The individual genealogical trees are the result of particular  
68 cell states, and the probabilistic transitions between them. Thus, the process of  
69 neurogenesis can be described in two complementary ways: The Cell Lineage Tree  
70 (CLT) that describes the genealogical relationship between the individual cells gen-  
71 erated during development; and the State Diagram (SD) that describes the possible  
72 states that cells may take, and the stochastic transitions between these states. The  
73 functional mechanism underlying these descriptions is the mitotic process and its  
74 interaction with the gene regulatory network (GRN). Our challenge is to estimate  
75 the distribution of CLTs; to identify their underlying states and transitions; and  
76 then to posit a biologically plausible generative mechanism for their occurrence.

77 The purpose of this paper is to show that even sparse phenotypic and cell  
78 lineage data can be used to infer fundamental properties of neurogenesis and its  
79 organization. We begin by using previously published data to derive a stochastic  
80 state transition model of cortical neurogenesis, and from this we implement an  
81 abstract gene network that carries out the stochastic process. We then use a  
82 simulation of physical cell growth and mitosis to demonstrate that this GRN is able  
83 to steer in a distributed manner the quantitative cellular dynamics of neocortical  
84 neurogenesis.

## 85 **4. Results**

### 86 *4.1. Cell lineage Trees*

87 The Cell Lineage Tree is an acyclic directed graph in the form of a rooted  
88 binary tree, in which the vertices represent physical cell instances, and the directed

edges represent the genealogical relationships between mothers and their daughter cells. The root of the tree is the earliest stem cell (neuroepithelial cells in this case); the internal nodes of the tree are dividing multipotent or pluripotent precursor cells; and its leaf nodes are non-dividing terminally differentiated cells (neurons and glial cells).

Measurements of lineage subtrees indicate that at least in vertebrates the lineage mechanism is stochastic rather than deterministic (He et al., 2012). Thus, vertebrate lineage trees form a distribution over possible genealogies. When two new cell instances are generated by mitosis, fate transitions occur between the precursor and its offspring. If the precursor divides symmetrically it will produce two daughters with identical cell fates, and thus identical phenotypes. However, if it divides asymmetrically, the precursor will produce two cells that inherit distinct gene expression products, and as a consequence may have different cell fates. In principle, we could measure the feature vector  $f$  over all cell instances. But such an exhaustive description is not yet technically feasible. Thus, for the present purposes, we assume that the feature vectors can be observed only over terminally differentiated cells. That is, we can observe and classify the phenotypes of terminal cells in terms of their neuronal morphology and behavior. Figure 2A shows a simple CLT, for purpose of explanation. The terminal states of this CLT are categorized into three types ( $A$ ,  $B$ ,  $C$ ) based on a set of features  $\{f_A, f_B, f_C\}$ , which we assume can be observed only in terminal cells.

#### 4.2. *Cell Lineage Trees for mouse cortical neurogenesis*

We obtained estimates of the distributions of terminal neuronal types in mouse area 3 and 6 from the work of Polleux et al. (1997a), who used pulse  $^3H$ -thymidine injections made throughout corticogenesis to measure the variation of cell cycle

114 duration, cell cycle exit probability and laminar fate as functions of developmental  
115 time. Following their data and methods, we computed the temporal generation  
116 of neuronal types by numerical solution of the continuous differential equations  
117 describing cell proliferation and differentiation (Polleux et al., 1997b) (Figure 1).  
118 We then used these population distributions together with a probability-generating  
119 function (Bremaud, 1988) to generate probabilistically instances of cortical cell  
120 lineages (Figure 1).

### 121 4.3. *State Diagrams*

122 An alternative view of neurogenesis is one that describes the underlying generic  
123 cell states and their transitions, rather than the genealogical relationships between  
124 particular cell instances. We will call this alternative view the State Diagram (SD).  
125 It is a weighted directed graph whose vertices represent cell states, and whose  
126 weighted edges represent the stochastic transitions between states that occur at  
127 cell mitosis. Whereas the CLT describes both terminal cell identities and their  
128 individual ontogenies, the SD explains the experimentally observed numbers and  
129 dynamics of production of neuronal types in terms of state transition probabilities.

130 The SD begins from an initial precursor cell state; for example, the state of  
131 a neuroepithelial cell. When a cell undergoes mitosis, it generates two daughter  
132 states that will themselves generate subtrees of states, until a terminal state is  
133 reached. Because the SD vertices are states and not specific cells, cells that have  
134 exactly the same state are represented by the same single vertex. The numbers  
135 of cell transitions between one state and a different one are accounted for in the  
136 probabilistic weights of the edges that join the states. However, the sum of the  
137 probabilities across all the possible transitions away from a mother state is 2 not 1,  
138 because always two daughter states must be generated.

139 The SD can have different degrees of resolution, according to the mapping  
140 of individual physical cells to their possible underlying cell states. Trivially, any  
141 collection of lineage trees can be encoded exhaustively by an SD in which each  
142 and every cell instance is assigned to its own unique state (Figure 2B). Although a  
143 high resolution representation of this type is easy to generate, the number of states  
144 increases exponentially with the complexity of the cell lineage trees. The SD soon  
145 becomes intractably large, and the number of unique states and transitions rapidly  
146 exceeds the amount genetic information available to encode it.

147 A more suitable mapping of cells onto states assumes that biological processes  
148 are often best explained by models with low but noisy dimensionality. This is  
149 likely true for cell lineages, where only a very small set of all possible internal  
150 genetic expression profiles are visited by cells during development (Kauffman and  
151 Kauffman, 1993), and because very similar cell division sequences occur across  
152 the distribution of all lineage trees. Such a reduced encoding involves collapsing  
153 high dimensional graphs into subgraphs that have the same or similar underlying  
154 states and transitions. The example SD (Figure 2C) shows the principle of this  
155 reduction of redundant subtrees. The result is a more compact representation that  
156 describes the same developmental process, but using fewer states.

157 The general problem is to find such a low dimensional SD that is still able  
158 to account for most of the variance in the experimental data. We approached  
159 this problem by spectral clustering (Chung, 1997; von Luxburg, 2007), a type  
160 of clustering algorithm that can be applied to graphs. Our goal was to obtain an  
161 appropriate embedding of the full dimensional SD into a similarity matrix, such  
162 that the pairwise distance between cell states in the embedding space reflects  
163 their similarities in terms of terminal cell types than those two states give rise to.

164 Once the full SD is embedded into an Euclidean space, simple algorithms such as  
165 hierarchical clustering can be used to cluster cell states into smaller subsets and  
166 thereby generate a lower dimensional, more easily interpretable SD representation  
167 of the cell lineage.

168 Since the SD states can be characterized by feature vectors, the reduced SD also  
169 models implicitly the statistical distributions over the feature profiles characteristic  
170 of each state, and the genealogical relationships between these feature states.  
171 Unfortunately we do not have data for the internal nodes of the SD (but see (Pfeiffer  
172 et al., 2016)). However, the feature vectors for the terminal states are known, and so  
173 we can estimate the feature profiles of the hidden vertices by propagating the known  
174 features backward into the hidden network. In this way the precursor states are  
175 mapped to corresponding linear combinations of terminal features. These profiles  
176 are a prediction of the contributions of the various precursors to the different  
177 final neuronal fates. For convenience we visualize these relationships by suitable  
178 coloring of the SD graph. The feature vectors of terminal states are associated  
179 with unique color vectors. These colors are then propagated backward into the  
180 network as proxies for features. The ‘colors’ of the precursor cells provide a visual  
181 impression of the fates to which they will contribute (Figure S2 and Figure S4).  
182 The SD states are an estimate of the hidden biological cell states  $S$ . For example,  
183 we may take this estimate to be  $f$ . And so each node of the SD is labeled with a  
184 vector whose elements correspond to experimentally observable features  $f_j$ , such  
185 as the expression of a particular set of genes, or morphological features.

#### 186 4.4. *State Diagrams for mouse cortical neurogenesis*

187 We used our spectral clustering method to estimate the SD underlying the  
188 development of cortical areas 3 and 6 of the mouse. The dynamics of cellular



189 division and differentiation during development of these areas have been quantified  
190 using the mitotic history technique, which selectively monitors the proliferative  
191 behavior of defined cohorts of precursor cells generated at particular time points  
192 (Polleux et al., 1997b; Dehay and Kennedy, 2007). However, the behavior of the  
193 individual lineage trees supporting these population dynamics is unknown. There-  
194 fore we reconstructed probable lineage trees by sampling from the experimentally  
195 determined cell distributions (Figure 1). While the topologies of these trees are  
196 stochastic, their overall distribution is constrained by the experimentally observed  
197 distribution over different terminal cell fates.

198 We analyzed 60 such reconstructed lineages from area 3 and 6 of the mouse  
199 cortex. These lineages contained a total of 3263 cell instances (1549 in area 3 and  
200 1714 in area 6). The terminal cells were labeled as either *Layer 6b* (L6b), *Layer 6a*  
201 (L6a), *Layer 5* (L5), *Layer 4* (L4), *Layer 2/3* (L2/3), or *Glia*. Precursor cells were  
202 labeled as *Unknown*. The complete, unreduced, SD was composed of 6 terminal  
203 states; with 765 unknown precursor states in area 3 and a further 848 unknown  
204 precursor states in area 6. Spectral clustering for both areas was performed on the  
205 combined dataset. The combination of data allows the method to exploit possible  
206 similarities between the SDs of the two areas (Figure 3).

207 The original data is fully described by a SD of 519 dimensions, in which each  
208 cell has a corresponding state. Similar states generate cells with identical fates,  
209 and so can be collapsed into a unique state leading to a reduced SD with only 10  
210 dimensions with negligible loss of accuracy. Models with even fewer dimensions  
211 are also able to describe the data, but with less accuracy. In order to compare  
212 the performance of SD models of different dimensions, we estimated the model  
213 error as the number of incorrectly generated terminal cells types over the total

214 number of cells produced at the end of the developmental process. This error was  
215 compared against that of a complementary scrambled model, obtained by random  
216 permutation of cell states.

217 The accuracy of the SD models for area 3 and 6 was assessed for the homoge-  
218 neous (HM), the non-homogeneous (NM) and the time-dependent (TM) Markov  
219 process. In the HM model, transition probabilities are independent of time, and  
220 so at low model dimensions the cell output distributions have long tails because  
221 of small state transition probabilities, which cause a small proportion of cells to  
222 undergo many rounds of division (Figure S6 and S7). Convergence to the target  
223 distribution occurs only after a great number of cell divisions, which is unrealistic  
224 for biological processes. We therefore introduced time dependence by applying  
225 age-dependent probability distributions in the NM model: Each state has unique  
226 outgoing transition probabilities, and a maximal number of possible self-replicative  
227 divisions. This assumption truncates the long tails of the HM approach, forcing  
228 cells to progress through the differentiation path. Finally, in the TM model, each  
229 transition probability is computed for each round of cell division. This model  
230 reproduces accurately the cell distributions as well as their temporal dynamics.  
231 However, this accuracy comes at the cost of a large number of parameters. By  
232 contrast, the HM model requires a large number of cell states for an accurate  
233 prediction. Both cortical areas are best described by the NM model, which is able  
234 to reproduce closely the system dynamics, and offers a good trade-off between  
235 model complexity (31 or 10 dimensions) and model accuracy (11% or 18% model  
236 error) (Figure 4A, B).

237 The NM 10 dimensional SD model explains 82% of the data, and is the  
238 most visually intuitive for reasoning over the logic underlying the developmental

processes of area 3 and 6. The black node (with centered white dot) represents an initial homogeneous population of precursor cells, which then divide into subpopulations of precursor cells having different neurogenic potentials. A small proportion of cells are fated very early on to develop exclusively toward granular (L4) or supragranular layers (L2/3); and a large pool of heterogeneous precursor cells are less fate restricted (Figure 4B). The 31 dimension SD model is more precise: It explains 89% of the data, but it is less intuitive. A striking difference of this model with respect to the 10 dimension SD case, is the presence of two distinct initial populations that develop differently according to their fate restriction (Figure 4A). It is noteworthy that the precursor pool has some degree of plasticity in the sense that many cell states have bidirectional transitions, as has been observed in the cortical lineages of primates (Betizeau et al., 2013).

The SD's above were computed over the combined lineage datasets for areas 3 and 6. However, we track the contributions of each dataset, and so it is straightforward to decompose the combined SD into the separate SDs describing each area (Figure S5). The reduced SDs for area 3 and 6 are strikingly similar (Figure 4C, D), suggesting that only minimal changes in a single model are sufficient to explain observed differences of neurogenesis in individual areas.

#### 4.5. Estimates of SD gene expression patterns

So far we have interpreted the SD in terms of its propagation of terminal cell fates that are largely morphological, e.g. L2/3 pyramidal cell. However, SD models can also be interpreted in the light of the underlying gene expression process. For example, one might choose for features  $\{f_1, f_2, \dots, f_n\}$  the real, observed transcription factor expression levels. Such data were not available to us at the beginning of this project. However, for illustration of the principle we used

264 calibrated gene expression levels in cortical neurons obtained from a transcriptome  
265 atlas of cortical layers in the adult mouse area 3 (Belgard et al., 2011). Of the 11411  
266 gene probes used in that atlas, we consider only the subset of 1751 transcription  
267 factors. We applied *k-means* clustering to this dataset and thereby identified 12  
268 clusters of transcription factors that have similar expression patterns across the  
269 cortical laminae (Table S1). Each lamina is associated with one of the terminal  
270 neuronal types, and so each neuronal type is associated with a characteristic  
271 distribution across the 12 transcription factor clusters. Because the clustering is  
272 based on adult expression data, the distributions of the feature vectors are known  
273 only for terminal cell fates. However, as described above, our spectral clustering  
274 method can be used to propagate the adult values backward into the lineages and  
275 thereby provide a prediction of the expected transcription factor profiles to be  
276 found in the various SD precursor states (Figure 5).

#### 277 4.6. Abstract Gene Regulatory Networks

278 The second, complementary model, is functional. The states and state tran-  
279 sitions are implemented implicitly by a *genotypic* model (or Gene Regulatory  
280 Network, GRN) (Figure S1C). In this case the interactions between genes and  
281 transcription factors are explicitly modeled. The network is designed in such a way  
282 that the global developmental process arises from the local dynamics of genes in  
283 individual cells. This model is visualized as a graph (not a tree), in which the nodes  
284 represent genes, and the edges represent interactions between genes. Importantly,  
285 the genotypic model is mechanistic in that it not only expresses allowable states  
286 and state transitions, but also declares the causal mechanisms by which the states  
287 are implemented, and reached.

#### 288 4.7. An abstract GRN for mouse cortical neurogenesis

289 We will describe in detail below how the State Diagram (SD) can be estimated  
 290 from experimental data, and how a GRN can be constructed that expresses this  
 291 SD (and therefore the observed experimental data). Briefly, we first show that a  
 292 low dimensional SD, composed of a small set of states, is sufficient to explain  
 293 the generation of the different morphological cell types of the neocortex. This  
 294 phenotypic model is then matched to a corresponding genotypic model. Because  
 295 this problem is ill-posed (multiple genotypic models are able to explain a single  
 296 phenotypic model), we restrict the domain of solutions by seeking a biologically  
 297 realistic model based on a GRN. In our implementation, division asymmetry leads  
 298 to differential inheritance of transcription factors in the daughter cells. This process  
 299 is used to drive changing rates of cell numbers and types produced.

300 The SD generative model derived above is an example of a *phenotypic model*  
 301 that describes the observed experimental data by assigning to each cell a state, and  
 302 probability of transitions between those states at the time of cell division. This is  
 303 essentially a phenomenological description of the statistics of neurogenesis. How-  
 304 ever, the question of the actual biological mechanism that expresses this statistical  
 305 behavior is a much deeper one. Biological systems do not have a single constructor  
 306 with global knowledge, able to direct all aspects of development. Instead, the only  
 307 construction information available resides in the genetic instructions present in, and  
 308 essentially localized to, each cell. The challenge then, is to implement the complex  
 309 process of biological development as a *genotypic model* of neurogenesis. In this  
 310 model developmental control is localized to gene regulation within individual cells  
 311 (Figure S1C). The result of the operation of the GRN, distributed in its various  
 312 configurations across all the lineages of neurogenesis, should be observable as the

SD. Thus, we need to make the bridge from gene-level dynamics in individual cells, to the population-level stochastic behavior of the SD.

We have previously reported a formal language able to describe cellular and molecular processes that support cortical development (Zubler and Douglas, 2009). In particular, that language is able to control the development of a simple laminated cortical column (Zubler et al., 2013). However, in that previous work the generation of different cell types required precise ad hoc tuning of a system of differential equations. By contrast, our goal here was to create a genetic network model based on observed cellular mechanisms that is robust to intrinsic noise, reliable in execution, and flexible in the range of cell types it can generate.

The cellular machinery is composed of several layers of regulation. At the outermost layer, functional proteins fulfill specialized tasks such as structural support, movement, and cell morphology. Deeper in the regulatory machinery, DNA-binding regulatory proteins (transcription factors), define the progression through different cell activity states by regulating the gene expression profile of each cell. Transcription factors influence one another's expression over time by binding to specific gene regulatory regions. The overall combination of the core regulatory network composed of transcriptions factors as well as the functional genes responsible for the cell phenotype, is referred to as a *Gene Regulatory Network* (GRN). However, the description below focuses largely on the transcriptional aspect of the GRN.

The concentration of each gene  $x_i$  is computed as a function of the concentration of other genes  $\mathbf{x} = x_1, x_2, \dots, x_n$  by the rate equation:

$$\dot{x}_i = k_1 \mathcal{F}_i(\mathbf{x}) - k_2 x_i \quad (1)$$

with:

$$\mathcal{F}_i(\mathbf{x}) = \sum_j^n \beta_{ij} \prod_j^n Z_{ij}(x_j) \quad (2)$$

334 The function  $\mathcal{F}_i(\mathbf{x})$ , or *sigma-pi function*, is a linear combination of elements  
 335  $Z_{ij}$ , each of which represents the binding of a transcription factor  $j$  on gene  $i$  as  
 336 a function of its concentration  $x_j$  according to a sigmoidal probability binding  
 337 function, the Hill function  $Z$ . Linear combinations of  $Z$  elements, determined by the  
 338 coefficients  $\beta_{ij} \in \{0, 1\}$ , describe how transcription factors interact with each other  
 339 by steric interactions. This formulation provides a model to express transcriptional  
 340 networks as compositions of continuous Boolean logic gates (Figure S8), for which  
 341 we propose an intuitive formal language based on logic gates.

342 Decisions leading to the acquisition of an appropriate cell fate rely on the ability  
 343 of cells to commit to different stable states. A system that can perform such a  
 344 task is a module with competitive and cooperative interactions. The most simple  
 345 example of such a system is the bistable switch (Niwa et al., 2005; Huang et al.,  
 346 2007), in which two auto-catalytic transcription factors  $A$  and  $B$  negatively regulate  
 347 each others expression:

$$\begin{aligned} a &= k_1 \text{AND}[\text{OR}[Z(a), \text{NOT}[Z(b)]], Z(I)] - k_2 a \\ b &= k_1 \text{AND}[\text{OR}[Z(b), \text{NOT}[Z(a)]], Z(I)] - k_2 b \end{aligned} \quad (3)$$

348 where  $a$  and  $b$  refer to the concentrations of the proteic product of genes  $A$   
 349 and  $B$ , and  $k_1 = 1$  and  $k_2 = 1$  represent production and degradation constants  
 350 respectively. The system can be driven toward a specific state by an input  $I$  and  
 351 is explicitly designed to display hysteric behavior upon input withdrawal: The  
 352 network can remember the existence of past input signals (Figure S9). This design  
 353 feature confers remarkable stability of the gene expression, and makes the dynamics

354 of the module dependent only on an initial input signal (Jacob and Monod, 1961;  
355 Glass and Kauffman, 1973; Hartwell et al., 1999).

356 Biological development can be viewed as a sequential progression of precursors  
357 through different gene expression profiles; each cell state is associated with a  
358 characteristic profile. Thus, each lineage tree expresses one stochastic lineage of  
359 profiles arising from a given root precursor. The crucial question for understanding  
360 the dynamics of neurogenesis is how distinct profiles arise during the mitoses of the  
361 lineage, and so allow different fates for daughter cells. In our model this important  
362 property is due to possible differential distribution of transcription factors to the  
363 daughters. Each gene  $X$  is characterized by an asymmetry constant parameter  $\alpha_X$ ,  
364 corresponding to the asymmetric division constant of its protein. Asymmetrical  
365 cell divisions lead to different distributions of transcription factors in the daughter  
366 cells, and thus to different gene expression profiles. Thus, cells regulated by a  
367 single bistable switch with asymmetry constants  $\alpha_A$  and  $\alpha_B$  can produce a range of  
368 cells with differing fates as a function of the division angle  $\omega$ , the orientation of the  
369 mitotic spindle with respect to the internal distribution of substances (Figure 8). We  
370 set the required  $\alpha$  for each substance in the bistable switch given a normalization  
371 constant  $N$ , such that  $-1 \leq \alpha_X \leq 1$ :

$$\begin{aligned}\alpha_A &= N \left( \frac{\sin(\omega)}{\cos(\omega) + \sin(\omega)} \right) \\ \alpha_B &= N \left( 1 - \frac{\sin(\omega)}{\cos(\omega) + \sin(\omega)} \right)\end{aligned}\tag{4}$$

372 Beginning with the initial state “0” with low expression of both genes  $A$  and  $B$   
373 (black cells), the activation of the input signal pushes cells to an undecided state  
374 “ $AB$ ” characterized by high levels of  $A$  and  $B$  expression (orange cells). Either by  
375 the presence of an external influence, or by asymmetric cell division, cells can  
376 jump to states “ $A$ ” or “ $B$ ”, where only one gene of the bistable switch dominates the



377 expression (pink or blue cells). Depending on the extent of the jump, each cell has  
378 a defined probability to reach new, otherwise inaccessible states. The irreversibility  
379 of jumps in the genetic landscape is implemented here as a dependency of the  
380 asymmetry constants on the gene product concentrations of the bistable genes.  
381 Once the motif reaches status “A” or “B”, further asymmetric division are inhibited,  
382 thereby limiting backward jumps to previous undifferentiated states.

383 The stochastic progression of precursors down differentiation paths can be  
384 modeled by a sequence of multiple genetic bistable switches, where each switch  
385 represents a branch in the differentiation decision tree and transition probabilities  
386 are mapped to cell division angle probabilities. Additional genes are required to  
387 detect specific transcription factor expression profiles and activate downstream  
388 functional programs. Control of precursor division is implemented by an inde-  
389 pendent clock mechanism that abstracts the complexities of the cell cycle and its  
390 phases. For simplicity it is assumed here to be a Gaussian distributed variable,  
391 independent on other events of the GRN. This basic genetic circuit is used to  
392 control cell fate decision at the moment of cell division, and to link the activa-  
393 tion of different functional genes, such as genes responsible for cell migration,  
394 differentiation or apoptosis.

#### 395 *4.8. Self-construction of a volume of cortex in Cx3D*

396 Finally, we validate the behavior of the GRN in a simulated physical environ-  
397 ment using Cortex3D (Cx3D) (Zubler and Douglas, 2009), an agent and Java based  
398 simulation environment for investigating the physical growth of multicellular struc-  
399 tures. This approach demonstrates the principles underlying the self-construction  
400 of a simple laminated cortical column and its neuronal connectivities (Zubler et al.,  
401 2013). In contrast to our earlier ad hoc system of differential equations for gene

402 regulation (Zubler et al., 2013), we propose here a formal genetic language to  
403 design biologically plausible gene regulatory networks. We go on to demonstrate  
404 that the derived genetic network is able to control the generation of cortical laminae  
405 for different cortical areas by intrinsic genetic specification and by the information  
406 provided by the environment.

407 For the design of the GRN, sequences of bistable genetic motifs are used  
408 to encode cell fate decision at division and implement a genetic version of the  
409 state diagram for area 3 and 6. The SD was enhanced to introduce states for the  
410 generation of additional cell types (L1, subplate, and glial precursors cells), and  
411 to further reduce the overlap in the production of different cell types in time, as  
412 this has a dramatic effect on the stability of the simulation and the generation of  
413 homogenous layers.

414 Each state in the SD is mapped to 2 genes whose interactions implement the  
415 required bistable behavior. In addition, these genes are coupled to members of other  
416 bistable switches, or possibly to functional genes that execute cellular behaviors  
417 (Figure 6). State transition probabilities are encoded in the mitotic division angles  
418 that control the stochastic distribution of symmetric and asymmetric cell divisions.  
419 The core transcriptional network regulating the asymmetric distribution of cell fate  
420 determinants is composed of 36 genes. Further 24 housekeeping genes decode  
421 transcriptional expression into function, such as cell differentiation, migration, and  
422 other behavioral outcomes.

423 The developmental model was then implemented in Cx3D (Figure 7). The  
424 simulation begins with an array of precursor cells in the neural epithelium lining  
425 the lateral ventricles (Figure 7, black cells). Each of these cell contains an identical  
426 copy of the genetic regulatory network (Figure 6A), initialized to its neuroepithelial

precursor configuration. The precursors are aligned on the apical surface, and this orientation is used to establish the cell internal polarity axes.

From this point onward, the behaviors of the distributed GRNs and the cells that they control are entirely autonomous. There is no intervention by a global controller, no explicit or global clock, and no explicit spatial coordinate frame. The only spatial cues are a pair of complementary morphogenic gradients in the medial/lateral axis of the neuroepithelial plate (Greig et al., 2013). The expression states of the distributed GRNs trigger their cells to undergo symmetrical or asymmetrical divisions according to their division angle, thereby forming the desired populations of successive precursors. The expression profiles at mitosis steer the stochastic transitions to successor states in the daughter cells. Mitosis is controlled by individual local cell cycle machines that induce cell cycle progression in precursors cells until they reach terminal differentiation. The entire process of neurogenesis from neuroepithelial cell to differentiated neurons involves some 20 mitotic divisions (Figure 6B).

Initially (E9-E12), the precursors progress through a sequence of increasing asymmetric divisions that lead to the production of the marginal zone (L1) and subplate cells, forming the early preplate. At the same time the VZ is formed. It is composed of radial glial cells (RGC) characterized by the extension of a radial process that often reaches the pial surface. Differentiating precursor cells that exit the cell cycle migrate along radial glial processes, constituting the successive waves of cell types that form the cortical plate in a inside-out manner. Migration is directed by local integration of guidance cues secreted by the marginal zone. A membrane bound stopping signal prevents cells from migrating past the pia. The density of cells in the marginal zone was also increased to provide physical

452 containment of upwardly migrating cells.

453 In a subsequent phase (E13-E16) a second germinal layer, the SVZ is formed.  
 454 In contrast to the VZ, precursor cells of this zone, the BPs, lose their radial process  
 455 and apical polarity. In our simulation, lost processes are not degraded and continue  
 456 to provide a scaffold along which neurons can migrate, increasing significantly  
 457 the stability of the formation of distinct laminae. In this second phase, granular  
 458 (L4) and supragranular (L2/3) are produced. The construction process ends with  
 459 the establishment of the cortical sheet, and a residual germinal layer composed of  
 460 glial cell precursors. Subsequently, corticogenesis would continue with a sequence  
 461 of symmetric division for the generation of glial cells, and the growth of the first  
 462 neural connectivities. These aspects are beyond the scope of the present paper,  
 463 which is concerned only with the general principles of the GRN and its derivation.

464 The simulation exhibits a clear arealization of laminar organization that con-  
 465 form to the characteristics of areas 3 and 6 (Figure 7). The percentages of various  
 466 neuronal types produced by the simulation in both areas also conform remarkably  
 467 well to experimental observation (Table 1). There is a short intermediate zone be-  
 468 tween these two areas, corresponding to a cytoarchitectural boarder. This transition  
 469 zone in the simulation may be analogous to area 4 that is interposed between areas  
 470 3 and 6 in mouse cortex, but which was not explicitly modeled.

471 In the simulation, areal specificity is cued by the initial gradient of morphogens  
 472 aligned with the medial/lateral axes of the developing sheet. The concentrations  
 473 of these morphogens are transcription factors for a gene pair ('g89A' and 'g89B',  
 474 Figure 6). These genes bias neurogenesis toward either an area 3 or an area  
 475 6 phenotype by slightly changing the distribution of the precursor pool, when  
 476 threshold conditions on the morphogen concentrations are satisfied. The 'g89' is

expressed on lineages leading towards L5 pyramidal cells. The onset occurs some 4 divisions before final differentiation, and there affects the relative generation of precursors fated towards layers 4/5. Thus, development towards area 3 or 6 occurs through a small bias in the distribution of precursor cells, localized to particular region of the lineage tree (and so a time window) well before differentiation (Figures S5, 6B).

## 5. Discussion

We use ‘self-construction’ to refer to the process whereby a system is able to make use of physically encoded rules to steer its own elaboration, without the intervention of any kind of external supervisor. By contrast, ‘development’ refers to the biological process whereby a single, or small number of precursors replicate and differentiate toward a very large, diverse population of differentiated and functionally organized cell types. Thus, questions of self-construction are concerned with the abstract principles that underlie development of biological systems, but might equally well be applied to a future technology.

We choose to study biological self-construction in the neocortex, because cortical development presents many interesting challenges. For example, cortical neurons are produced far from their final location in the adult and so must undergo a long migration before they can complete their differentiation and formation complex long-distance connections. Further, the cortical construction process results in a rather uniform laminar sheet on which is superimposed a more detailed structural and functional arealization, suggesting that subtle modifications of a general process of neurogenesis may be sufficient to explain the apparent complexity of cortical neural circuits.

501 Cortical cytoarchitecture and its parcellation into distinct areas reflects the  
 502 spatiotemporal modulation of neurogenesis (Dehay et al., 1993; Polleux et al.,  
 503 1997a; Dehay and Kennedy, 2007; Rakic, 2009). From its simple origins as a single  
 504 layer of proliferative cells in the embryonic dorsal ectoderm, the cortex grows  
 505 through self-replication of a small population of precursor cells. The interplay  
 506 between these many local mechanisms of cellular interaction, and their relationship  
 507 to global system behavior, are easier to grasp through detailed models and their  
 508 simulations (Fisher and Henzinger, 2007).

509 Here we have used a modeling approach to address the question of how a single  
 510 cellular regulatory system could determine the generation of a diversity of neurons,  
 511 including their laminar location. Of course, sufficiently detailed data describing  
 512 the full mechanism of gene regulation and its consequences for the behavior of  
 513 individual precursors underlying development are not yet available. However, we  
 514 demonstrate here that it is possible to obtain substantial insight into developmental  
 515 mechanisms using only sparse experimental data. With less than 40 genes we are  
 516 able to recapitulate the steps of cortical development in silico with Cx3D.

517 Our approach has two phases. In the first phase the experimental data describing  
 518 the generation of various neuronal types is used to estimate the stochastic SD  
 519 governing the generation of possible cell lineage trees (*phenotypic model*). Then,  
 520 in the second phase we implement the SD with a compact GRN-like state model  
 521 (*genotypic model*) whose behavior then satisfies the experimentally observed  
 522 dynamics of neurogenesis with quantitatively very similar cell distributions. This  
 523 GRN is composed of abstract genes, whose patterns of expression determine the  
 524 observed range of cell behavior.

## 525 5.1. *State model of cortical neurogenesis*

526 Hidden Markov Trees, which model Markov Tree processes over a set of trees of  
527 observed variables, and their conditional dependencies, have been used successfully  
528 to cluster cells and infer cell states from partial lineage tree reconstructions (Olariu  
529 et al., 2009; Pfeiffer et al., 2016). However, such inference requires a relatively  
530 large amount of data and is impractical for very sparse samples unless there are  
531 additional constraints on the probability distributions. Instead, we derived a lower  
532 dimensional representation of lineages using a simpler approach based on spectral  
533 clustering on graphs, whereby it is possible to exploit lineage information to cluster  
534 cells according to their phenotype, and that of their daughters.

535 We have introduced the concept of a SD to capture the complexity of the cell  
536 lineages. The SD model assumes that the underlying biological mechanisms can be  
537 modeled as a Markov process, according to which each cell, with its characteristic  
538 features, can be completely described by an unobserved state. The evolution of  
539 cell states is defined by the cell's current state, which comprises the cell's internal  
540 state and its immediate surroundings. In contrast to our related work (Pfeiffer et al.,  
541 2016) in which phenomenological data is used to classify progenitors cells in the  
542 primate cortex, we address here the use of genetic markers (transcription factors)  
543 to infer the probable developmental pathways followed by precursor cells until  
544 their terminal differentiation during murine corticogenesis.

545 Because we have only sparse data (i.e. we observe gene expression profiles on  
546 terminal cells only), we have used a simple approach based on spectral clustering,  
547 by which we cluster potential cell states according to the distributions of cell  
548 types that they are able to generate. The method was applied on cortical lineages  
549 inferred from experimental developmental data for areas 3 and 6. By this method

we obtained a low dimensional age-dependent model that explains neurogenesis in both cortical areas, and which, in contrast to homogeneous Markov processes is able to explain this developmental process using only a restricted number of states and parameters.

The SD model predicts that already at the neuroepithelial stage the precursor pool may be somewhat heterogeneous in terms of their fate potential. For example multipotent progenitor cells may coexist with a more specific population of cell fate restricted cells, as suggested experimentally (Franco et al., 2012; Guo et al., 2013). Interestingly, because transitions in our model are stochastic, progenitors may exhibit some plasticity, including the limited ability to revert to less differentiated states. Such transitions have been observed recently in primate corticogenesis, but have not yet been observed in the rodent cortex (Betizeau et al., 2013).

Surprisingly, the models for adjacent areas display many similarities and few significant differences. Key parameters in a single GRN distinguish the specification of cortical areas 3 versus 6. This observation suggests the presence of *genetic control points*, that is a small set of genes whose expression is able to control the switch between alternative cortical developmental programs. This finding agrees with the observed molecular similarity reported in neighbouring areas of the human frontal cortex (Johnson et al., 2009). More generally, this property suggests that the many areas of cortex within a species, could be affected by the settings of a small number of parameters in an otherwise rather generic control structure in accordance with biological observations (Ng et al., 2009; Bernard et al., 2012; Hawrylycz et al., 2012). This discovery poses the questions whether the emergence in the evolution of the primate neocortex is also due to changes in few, key genes, which lead to the generation of a much complex and diversified cerebral cortex,



575 and the significance of control points in biological processes in general (Dehay  
576 et al., 2015; Florio et al., 2015, 2016; Fiddes et al., 2018; Mitchell and Silver, 2018;  
577 Suzuki et al., 2018).

578 Obviously, the quality of the model depends strongly on the initial experimental  
579 classification of differentiated cell types, and a more extensive collection of data  
580 are required for a more precise version. In order to establish the general concept  
581 presented in this paper, we have relied heavily on the published cell birthdating data  
582 following pulse  $^3H$ -thymidine injections made throughout murine corticogenesis  
583 (Polleux et al., 1997a). However the same principles can be readily applied to gene  
584 expression (e.g. Figure 5) and other phenotypic data (e.g. (Pfeiffer et al., 2016))  
585 in future. While the recording in parallel of cell lineages and associated genetic  
586 markers is still a challenging technical endeavour, single cell tracking (Amat and  
587 Keller, 2013; Beattie and Hippenmeyer, 2017) or single cell profiling technologies  
588 (Bendall et al., 2014) would provide data at the necessary level of resolution.

## 589 5.2. *Gene regulation by asymmetrical division*

590 Our stochastic model of neurogenesis requires a number of distinct cell states in  
591 order to satisfy at least the experimental observations on which the model is based.  
592 The method of estimation of these states is constrained by additional more general  
593 structural knowledge such as the existence of lineage trees, binary mitosis, terminal  
594 states, etc. It is for this reason that it is possible to circumvent the seemingly  
595 ill-posed nature of moving from sparse data to an elaborate dynamical system that  
596 not only generates the original data, but will likely generalize to entirely different  
597 kinds of developmental data (e.g. gene expression, Figure 5).

598 The State Diagram alone provides a mathematical description of neurogenesis.  
599 However, it is difficult to relate that level of description to a biological mechanism.

600 The most interesting and experimentally useful aspect of this paper is the recog-  
 601 nition that it is possible to *implement* the global dynamics of a state model with  
 602 plausible biological mechanisms that have implications for further experimental  
 603 exploration. The implementation is based on basic cellular processes such as gene  
 604 regulation, cell division, and asymmetrical repartition of cellular components. In  
 605 particular, the importance of planar segregation of fate determinants during cortical  
 606 developmental processes has been recognized experimentally (Noctor et al., 2008).

607 We employ the concept of genetic regulation using a gene network design based  
 608 on small modules composed of bistable switches, each acting as an independent  
 609 functional component. The importance of multi-stability and modular organization  
 610 in molecular and genetic control has been recognized for over half a century  
 611 (Delbrück, 1949; Jacob and Monod, 1961; Glass and Kauffman, 1973; Hartwell  
 612 et al., 1999; Alon, 2006), however the modular networks reported here are arguably  
 613 the largest such systems yet, that have been configured to control the development  
 614 of complex tissue. We were surprised to find that the design of the GRN was less  
 615 difficult than we had anticipated. Because the individual modules are functionally  
 616 independent and self-restoring in their behavior, the interconnections between  
 617 modules are rather insensitive to parameter settings. The overall network inside a  
 618 given cell will converge toward its stable state, and it will finally trigger a mitotic  
 619 division, through which it copies itself to its offspring. Thus reliable modules  
 620 generate, by means of stochastic asymmetrical divisions, the desired distribution  
 621 of cells over neuronal types. In this way, even an homogeneous pool of precursors  
 622 can lead to the generation of diverse cell types. That is, the control of cell type and  
 623 numbers is implicit in the asymmetric distribution of gene products, and how the  
 624 genes influence one another's expression.

625 Currently, the model GRN is composed of arbitrarily named abstract genes.  
626 Their significance rests only in that this set and their interactions are necessary to  
627 satisfy the expression states and transitions required to control the developmental  
628 process. The relationship between those model genes and actual experimentally  
629 named genes expressed in particular developmental systems needs to be compre-  
630 hensively established. Establishing these relationships, as we have demonstrated by  
631 predicting the activation of transcription factors in the pool of precursor cells, and  
632 improving the model using the informative gene expression atlases will provide  
633 fruitful avenues for future research.

### 634 *5.3. Simulation of cortical neurogenesis*

635 The performance of the GRN was verified by simulation of neurogenesis using  
636 Cx3D (Zubler and Douglas, 2009). Cx3D respects physical processes such as mi-  
637 tosis, cell-cell interactions, movement and chemical diffusion in three-dimensional  
638 space. Each cell is an autonomous agent exerting only local actions, and using only  
639 locally available information. The physical behaviors of the cells are determined  
640 by the intracellular molecular processes expressed by the GRN. This large scale  
641 simulation of the physical mechanism makes it possible to bridge the scale between  
642 molecular processes and cell behavior.

643 The GRN is inserted into neuroepithelial prtecursor cells and initialized to a  
644 unique starting state. Each neuroepithelial cell contains also a simple cell clock  
645 that forces cells to divide at regular time intervals. Although the cell cycle length,  
646 in particular the length of the G1-phase, is correlated with the mode of cell division  
647 (Dehay and Kennedy, 2007; Pilaz et al., 2009; Lange et al., 2009; Arai et al.,  
648 2011) it was modeled here as an independent mechanism as the biological detail of  
649 this correlation is still unclear. The GRNs then orchestrate through their various

650 stochastic expressions in the successively generated cells, different molecular and  
651 physical processes leading to cortical lamination. It is by virtue of asymmetrical  
652 division that progenitor cells undergo progressive cell fate restriction in accordance  
653 with experimental observations (Shen et al., 2006; Gaspard et al., 2008).

654 Modulation of only a single gene was sufficient to steer neurogenesis towards  
655 the characteristic architectures of either area 3 or 6. This finding suggests a generic  
656 developmental program for corticogenesis across the cortex, where a few localized  
657 factors elicit the differences in neuron number that characterize cortical areas. This  
658 locally modifiable generic program could account for the multiplicity of cortical  
659 areas, despite a relatively restricted number of transcription factor gradients in  
660 the early forebrain (O’Leary et al., 2007; Sur and Rubenstein, 2005; Greig et al.,  
661 2013). During evolution there is a progressive increase in the number of cortical  
662 areas reaching as many as 140 in macaque (Essen et al., 2011), despite an expected  
663 conservation of the early patterning of the forebrain (Donoghue and Rakic, 1999;  
664 Rash and Grove, 2006; Monuki and Walsh, 2001; Bayatti et al., 2008; Šestan et al.,  
665 2001; Sur and Rubenstein, 2005). It is likely that such a generic developmental  
666 program can be spatiotemporally modulated by extrinsic factors including afferent  
667 fibers originating from the sensory periphery as shown experimentally (Dehay et al.,  
668 1996; Dehay and Kennedy, 2009; Rakic et al., 2009; Krubitzer and Kaas, 2005),  
669 which coupled to genetic changes could lead to diverse evolutionary scenarios  
670 (Striedter, 2005).

671 We have shown in this paper that sparse phenotypic and cell lineage data can  
672 be used to derive an abstract GRN whose dynamics are able to control the detailed,  
673 quantitative, neurogenesis of the areas from which the original data was obtained.

674 The remarkable reliability of the modeled neurogenesis rests in the multi-stable

675 and modular architecture of the GRN. Although mitosis may create offspring with  
676 different initial conditions, they will each reliably converge towards a permitted  
677 gene expression state and so to a recognizable precursor type of the cell lineage.  
678 Subtle and localized changes induced by mitosis in the stochastic distribution of  
679 transcription factors across offspring, can steer the overall profile of differentiated  
680 cells and their laminar location. The model can be used to explore and predict  
681 the forms of lineage and the resultant precursor pool sizes and relationships that  
682 precede the final adult cortical architecture.

683 While the present model of cortical neurogenesis is only an approximation to  
684 vast biological detail, it starts to explain the nature of the global coherence amongst  
685 multiple, distributed, locally independent cellular agents; and provides a useful  
686 tool for exploring the complex relationship between individual cell gene expression  
687 and population behavior underlying the development of the brain. Additionally it  
688 will also be a valuable tool for explaining diseases associated with gene regulation  
689 during cortical development.

## 690 **6. Acknowledgments**

691 We acknowledge helpful discussions with our SECO collaborators, in particular  
692 Kevan Martin, Christoph von der Malsburg, Michel Pfeiffer, and Adrian Whatley.  
693 This work was supported by European Union project grant FP7-216593 “SECO”.  
694 This work was also supported by LabEx CORTEX (ANR-11-LABX-0042)-HK,  
695 CD, and LABEX DEVweCAN (ANR-10-LABX- 061)-CD of Université de Lyon,  
696 within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated  
697 by the French National Research Agency (ANR); ANR-14-CE13-0036 (Primacor)  
698 and Fondation pour la Recherche Médicale (Equipe DEQ20160334943)-CD.

## 7. Methods

### 7.1. Cortical cell lineages reconstruction

We used published cell birthdate data from somatosensory cortex (Polleux et al., 1997a) to estimate the distribution of lineage trees underlying the neurogenesis of mouse area 3 and 6. Polleux et al. (1997a) employed pulse  $^3H$ -thymidine injections made throughout corticogenesis to measure the variation of cell cycle duration, cell cycle exit probability  $k_Q(t)$ , and laminar fate  $k_{QX}(t)$  as functions of developmental time  $t$ . Following their data and model we computed the temporal generation of neuronal types by numerical solution of the continuous differential equations describing cell proliferation and differentiation (Polleux et al., 1997b). We used these population distributions across developmental time to generate probabilistically instances of cortical cell lineage trees (Figure 1).

Cell proliferation can be seen as a discrete branching process whose time step  $\Delta t$  is equal to the cell cycle length. At each time step, cells either differentiate terminally with probability  $p_1 = k_Q(t)$ , or they divide with probability  $p_2 = (1 - k_Q(t))$  to form two daughter cells. These possibilities can be represented formally by the probability-generating function (pgf) (Bremaud, 1988):

$$f(s) = \sum_i^2 p_i s^i = k_Q(t)s + (1 - k_Q(t))s^2 \quad (5)$$

where  $p_i$  is the probability that a cell gives  $i$  offspring in the next generation and  $s^i$  is a dummy variable that accounts for the different numbers of cells generated. The pgf enumerates all the possible outcomes after one time step, and has the property  $\sum_i p_i = 1$ . We used this formula recursively to generate possible sequences of cells from single precursor cells. Sixty probabilistic lineage trees were computed for each of the two areas.

## 7.2. Graphical representation of the State Diagram

The State Diagram (SD) describes the states of cells that appear in the CLT, and the genealogical relationship between these states. For each state there is a corresponding vector of observed features  $\langle f_1, f_2, \dots, f_L \rangle$ . States for which features have been observed experimentally are defined as labeled, otherwise the states are unlabeled or hidden. We assumed that observed features (e.g. neuronal morphologies, gene expression) are available only for terminal cell states, and that the features of all the precursors are hidden.

It is convenient to represent the State Diagram in the form of a directed graph. Recall that  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  is a directed graph with vertices  $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$  and directed edges  $\mathcal{E} = \{e_{ij}\} \subseteq \mathcal{V} \times \mathcal{V}$ . In a weighted graph, each edge is assigned a specific value, its weight. For such weighted directed graphs, there is an asymmetric, non-negative adjacency matrix  $\mathbf{W}$  that associates each edge with a weight as following:  $w_{ij} = 1$  if there is a direct link that connects node  $i$  to node  $j$  or  $w_{ij} = 0$  otherwise. Also, we define the *in-degree* matrix  $D_{in}$  as the diagonal matrix of the sum of weights on incoming edges and the *out-degree* matrix  $D_{out}$  as the diagonal matrix of the sum of weights on outgoing edges:

$$D_{in}(j, j) = \sum_i w_{ij}, D_{out}(i, i) = \sum_j w_{ij} \quad (6)$$

Given a directed weighted graph, there is a natural random walk on the graph defined by a transition probability matrix  $\mathbf{P}$ , where  $p_{ij} = w_{ij}/d_{out}(i)$  for all edges, and 0 otherwise. Thus, in this naive random case, transitions on the outgoing edges are equally probable, and sum to 1. The situation for the State Diagram is somewhat different. Each vertex  $V$  of the State Diagram corresponds to a cell state, and each edge  $E$  asserts a genealogical relationship between connected states. Now

the transition probability matrix  $P$  represents the strength of these genealogical paths between states. That is, it represents the proportion of cells in the source state that will undergo each of the allowable transitions, multiplied by 2 to account for the doubling of cell number by mitotic division.  $P$  must be estimated from data.

### 7.3. Dimensionality reduction of the State Diagram

Given an SD and vectors of observed features  $\langle f_1, f_2, \dots, f_L \rangle$  for its labeled terminal nodes, we consider the task of computing a pairwise similarity measure between all nodes of the SD based on how unlabeled nodes are connected to labeled nodes. For undirected graphs, a widely used method for computing structural similarity is spectral clustering (Chung, 1997; von Luxburg, 2007). This method makes use of the spectrum (eigenvalues) of a similarity matrix to cluster data into groups of highly similar nodes. For our case of directed graphs, we introduce an approach based on the Laplacian  $\mathbf{L}$  of the normalized directed matrix:

$$\mathbf{L} = \mathbf{I} - \mathbf{D}_{out}^{-1} \mathbf{P} \mathbf{D}_{in} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (7)$$

where  $P$  is the directed transition probability matrix,  $\mathbf{D}_{out}$  is the out-degree matrix,  $\mathbf{D}_{in}$  is the in-degree matrix, and  $\mathbf{I}$  is the identity matrix.  $\mathbf{\Lambda} = \text{diag}[\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n]$  is the diagonal matrix of eigenvalues, and  $\mathbf{U} = [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_n]$  is the orthonormal matrix with eigenvectors of  $\mathbf{L}$  in each column.  $\mathbf{U} : \mathcal{V} \rightarrow \mathbb{R}^n$  provides an embedding of each vertex in an  $n$ -dimensional metric space. Each column of  $\mathbf{U}$  corresponds to an axis of the space, while each row of corresponds to the coordinates of a vertex in that space. The Euclidean distance  $\delta$  between pairs of nodes  $(r, s)$  provides a distance matrix:

$$\delta_{rs}^2 = (\mathbf{f}_r - \mathbf{f}_s)(\mathbf{f}_r - \mathbf{f}_s)^T \quad (8)$$



Mapping of the State Diagram to a  $n$ -dimensional space is particularly useful, because conventional algorithms such as hierarchical clustering can be applied there. We used the single linkage algorithm to perform clustering on the distance measure. Nodes whose distance was less than a specified threshold were clustered into a single node, which was assigned the average of their transition probabilities. The projection is in Euclidean space and so the feature vectors for each clustered node can be computed by solving a linear equation, because we assume that each node can be represented by a linear combination of feature vectors:

$$\mathcal{F} = \mathbf{U}\mathbf{F} \quad (9)$$

where  $\mathbf{F}$  is a  $n \times l$  matrix containing the features of the observed states,  $\mathbf{U}$  is a  $n \times n$  matrix, and  $\mathcal{F}$  is a  $n \times l$  matrix with observed and estimated features. For visualization purposes, each terminal state was also matched to a 3-element feature vector  $\mathbf{F}_{RGB}$  representing a unique color, and colors of all states were estimated by  $\mathcal{F}_{RGB} = \mathbf{U}\mathbf{F}_{RGB}$ .

We validated our spectral clustering method by measuring its performance on a set of artificial lineages generated by ‘ground truth’ models. The classification of cells to states by the algorithm was compared against 100 deterministic, stochastic and random cell lineages each composed of 5 states. The fraction of states misclassified by the algorithm are shown in the confusion matrices of Figure S3. The columns of the matrices represent instances of predicted states, while the rows represent instances of ground truth states. We found that deterministic ground truth models are recovered in 100% of cases, while probabilistic ground truth models are recovered in 80%. This decrease in performance on probabilistic models is due to misclassification of states as well as to the existence of multiple equally likely

789 solutions. The chance of random prediction of 5 states is estimated at 18%. These  
790 results demonstrate that a low dimensional SD can indeed capture the statistical  
791 variation of the cell lineage data at above chance level.

#### 792 7.4. Multi-type Markov Branching Process

793 A State Diagram can be interpreted as a Markov branching process with mul-  
794 tiple states. A branching process is a discrete-time random process that models  
795 a population in which each particle in generation  $t$  produces some number of  
796 individuals in generation  $t + 1$ , each of which can assume one of  $m$  different states.

797 Let  $S$  denote a finite set of states  $S = \{s_1, s_2, \dots, s_m\}$ , and  $Z_n = (z_1, z_2, \dots, z_m)$   
798 the vector of variables describing the population size at the  $n$ 'th generation in each  
799 state. Then the time-invariant transition probability  $p_{ij}$  describes the probability  
800 that a particle will transit from state  $i$  to state  $j$  (Markov property):

$$p_{ij} = \mathbb{P}(Z_{n,j} = z_j | Z_{n-1,i} = z_i) \quad (10)$$

801 The system evolution is completely characterized by the set of states, the  
802 marginal distribution of its initial state  $Z_0$ , and the transition probabilities between  
803 states. We write the joint probability distribution of  $Z_n$ :

$$\mathbb{P}(Z_n) = \mathbb{P}(Z_0) \prod_{t=1}^n \mathbb{P}(Z_t | Z_{t-1}) \quad (11)$$

804 By setting the elements of the weight matrix  $P$  equal to the probability of mov-  
805 ing from state  $i$  to a state  $j$ , the equation may be rewritten in matrix representation:

$$\mathbb{P}(Z_n) = \mathbb{P}(Z_0) \prod_{t=1}^n \mathbb{P}(Z_t | Z_{t-1}) = Z_0 \mathbf{P}^n \quad (12)$$

Markov models have limited ability to describe complex time-dependent processes using only a restricted set of states. Therefore, we extended this homogeneous Markov model (HM, probability  $\mathbf{P}$ ) by two further approaches. First, as a non-homogeneous model (NM, age-dependent probability  $\mathbf{P}(a)$ ). Here each state transition probability is multiplied with an additional parameter that is set to 0 once a maximal number of self-replicating divisions is reached. This has the effect of truncating the long tails that are characteristic of Markovian processes. Second, as a time-dependent model (TM, time-dependent probability  $\mathbf{P}(t)$ ) that explicitly encodes the state transition probabilities for each time point. In order to compare branching processes for these three different approaches and different model dimensions, we computed their errors as the number of misclassified cells (cells in wrong terminal states) over the total number of cells produced at the end of the developmental process.

### 7.5. Formal genetic language definition

We designed a genetic “language” in order to describe gene regulatory networks (GRNs). This language was based on a set of variables  $x \in \mathbb{R}_{\leq 0}$  that represent substance concentrations, and a set of allowed operations on the substance concentration values. This formalism greatly simplifies the construction of GRNs for developing systems as it is based on the design of the network topology, so that parameter tuning is reduced to a minimum. Although abstract, the formalism can be cast directly into the corresponding kinetic differential equations:

**Read.** Information about transcription factor concentrations is obtained from the environment through the Hill function  $Z$ , which computes the binding probability of a transcription factor to a promoter region given affinity constant  $\theta$ , cooperativity  $m$  and binding bias  $b$ .

$$Z(x + b, \theta, m) = \frac{(x + b)^m}{\theta^m + (x + b)^m} \quad (13)$$

831 **Write.** Information can be written to the environment by the production of  
832 a given substance according to the rate equation, which influences the current  
833 substance concentration.  $\mathcal{F}$  takes the form of one of the possible logic operations,  
834 or combinations thereof.

$$\dot{x} = k_1 \mathcal{F}[Z(\mathbf{x})] - k_2 x \quad (14)$$

835 **Distribute.** Information is encapsulated by the cell membrane, which prevents  
836 external agents from directly interacting/modifying the cellular molecular com-  
837 ponents, and so provides a protected environment in which the cell performs its  
838 local computation. During development, a cell  $c$  divides and distributes its internal  
839 components asymmetrically to daughter cells  $2c$  and  $2c + 1$ .

$$x_{2c} = x_c + \alpha x_c \quad (15)$$

$$x_{2c+1} = x_c - \alpha x_c$$

840 **Logic operations.** Logic operations are used to compute the result of the  
841 binding of multiple transcription factors to the promoter region, where  $y$ 's can be  
842 either the output of  $Z$  or the output of another logic operation.

$$\text{AND}(y_1, y_2) = y_1 \cdot y_2 \quad (16)$$

$$\text{OR}(y_1, y_2) = y_1 + y_2 - \text{AND}(y_1, y_2) \quad (17)$$

$$\text{NOT}(y) = 1 - y \quad (18)$$

843 **Derived logic operations.** The elementary operations can be composed into  
844 derived operations, for example:

$$\text{XOR}(y_1, y_2) = \text{AND}(\text{NOT}(\text{AND}(y_1, y_2)), \text{OR}(y_1, y_2)) \quad (19)$$

$$\text{NAND}(y_1, y_2) = \text{NOT}(\text{AND}(y_1, y_2)) \quad (20)$$

$$\text{NOR}(y_1, y_2) = \text{NOT}(\text{OR}(y_1, y_2)) \quad (21)$$

$$\text{NXOR}(y_1, y_2) = \text{NOT}(\text{XOR}(y_1, y_2)) \quad (22)$$

$$\text{TRUE}(y) = \text{AND}(y, y) \quad (23)$$

$$\text{FALSE}(y) = \text{NOT}(\text{TRUE}(y)) \quad (24)$$

845 Another useful derived operation is the threshold function  $Z_o$ , that indicates a  
846 threshold at any desired value  $tr \in [0, 1]$ :

$$Z_o(y, tr, \theta, m \rightarrow \infty) = Z(y + \theta - tr, \theta, m \rightarrow \infty) \quad (25)$$

847 Notice that for co-operativity  $m \rightarrow \infty$ , values of  $x$  are bounded to the set  $\{0, 1\}$ ,  
848 logic operations behave as Boolean logic gates, and the genetic language reduces  
849 to conventional Boolean algebra.

## 850 7.6. Software

851 Spectral clustering was implemented in Matlab R2012a. Graph visualizations  
852 were performed using a Cytoscape 3.0 plugin (DynNetwork). Cortical simulations  
853 were performed using Cortex3D (Cx3D) (Zubler and Douglas, 2009).

## References

- Alon, U. (2006). *An introduction to systems biology: design principles of biological circuits*. CRC press.
- Amat, F. and Keller, P. J. (2013). Towards comprehensive cell lineage reconstructions in complex organisms using light-sheet microscopy. *Development, Growth & Differentiation*, 55(4):563–578.
- Anthony, T. E., Klein, C., Fishell, G., and Heintz, N. (2004). Radial glia serve as neuronal progenitors in all regions of the central nervous system. *Neuron*, 41(6):881–890.
- Arai, Y., Pulvers, J. N., Haffner, C., Schilling, B., Nüsslein, I., Calegari, F., and Huttner, W. B. (2011). Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nature Communications*, 2:154.
- Bayatti, N., Moss, J. A., Sun, L., Ambrose, P., Ward, J. F. H., Lindsay, S., and Clowry, G. J. (2008). A Molecular Neuroanatomical Study of the Developing Human Neocortex from 8 to 17 Postconceptional Weeks Revealing the Early Differentiation of the Subplate and Subventricular Zone. *Cerebral Cortex*, 18(7):1536–1548.
- Beattie, R. and Hippenmeyer, S. (2017). Mechanisms of radial glia progenitor cell lineage progression. *FEBS letters*, 591(24):3993–4008.
- Belgard, T. G., Marques, A. C., Oliver, P. L., Abaan, H. O., Sirey, T. M., Hoerder-Suabedissen, A., García-Moreno, F., Molnár, Z., Margulies, E. H., and Ponting, C. P. (2011). A transcriptomic atlas of mouse neocortical layers. *Neuron*, 71(4):605–616.

- 877 Bendall, S. C., Davis, K. L., Amir, E.-A. D., Tadmor, M. D., Simonds, E. F., Chen,  
878 T. J., Shenfeld, D. K., Nolan, G. P., and Pe'er, D. (2014). Single-cell trajectory  
879 detection uncovers progression and regulatory coordination in human B cell  
880 development. *Cell*, 157(3):714–725.
- 881 Bernard, A., Lubbers, L. S., Tanis, K. Q., Luo, R., Podtelezhnikov, A. A., Finney,  
882 E. M., McWhorter, M. M., Serikawa, K., Lemon, T., Morgan, R., Copeland,  
883 C., Smith, K., Cullen, V., Davis-Turak, J., Lee, C.-K., Sunkin, S. M., Loboda,  
884 A. P., Levine, D. M., Stone, D. J., Hawrylycz, M. J., Roberts, C. J., Jones, A. R.,  
885 Geschwind, D. H., and Lein, E. S. (2012). Transcriptional Architecture of the  
886 Primate Neocortex. *Neuron*, 73(6):1083–1099.
- 887 Betizeau, M., Cortay, V., Patti, D., Pfister, S., Gautier, E., Bellemin-Ménard, A.,  
888 Afanassieff, M., Huissoud, C., Douglas, R. J., Kennedy, H., and Dehay, C.  
889 (2013). Precursor Diversity and Complexity of Lineage Relationships in the  
890 Outer Subventricular Zone of the Primate. *Neuron*, 80(2):442–457.
- 891 Bremaud, P. (1988). *An introduction to discrete probabilistic modelling*. Springer.
- 892 Chung, F. (1997). *Spectral Graph Theory*, volume 92 of *CBMS Regional Confer-*  
893 *ence Series in Mathematics*. American Mathematical Society, Conference Board  
894 of Mathematical Sciences.
- 895 Cárdenas, A., Villalba, A., de Juan Romero, C., Picó, E., Kyrousi, C., Tzika, A. C.,  
896 Tessier-Lavigne, M., Ma, L., Drukker, M., Cappello, S., and Borrell, V. (2018).  
897 Evolution of Cortical Neurogenesis in Amniotes Controlled by Robo Signaling  
898 Levels. *Cell*.

- 899 Dehay, C., Giroud, P., Berland, M., Killackey, H., and Kennedy, H. (1996). Contri-  
900 bution of thalamic input to the specification of cytoarchitectonic cortical fields  
901 in the primate: effects of bilateral enucleation in the fetal monkey on the bound-  
902 aries, dimensions, and gyrification of striate and extrastriate cortex. *J Comp*  
903 *Neurol*, 367(1):70–89.
- 904 Dehay, C., Giroud, P., Berland, M., Smart, I., and Kennedy, H. (1993). Modulation  
905 of the cell cycle contributes to the parcellation of the primate visual cortex.  
906 *Nature*, 366(6454):464–466.
- 907 Dehay, C. and Kennedy, H. (2007). Cell-cycle control and cortical development.  
908 *Nature Reviews Neuroscience*, 8(6):438–450.
- 909 Dehay, C. and Kennedy, H. (2009). Transcriptional Regulation and Alternative  
910 Splicing Make for Better Brains. *Neuron*, 62(4):455–457.
- 911 Dehay, C., Kennedy, H., and Kosik, K. S. (2015). The outer subventricular zone  
912 and primate-specific cortical complexification. *Neuron*, 85(4):683–694.
- 913 Delbrück, M. (1949). *A physicist looks at biology*. Connecticut Academy of Arts  
914 and Sciences.
- 915 Donoghue, M. J. and Rakic, P. (1999). Molecular Gradients and Compartments in  
916 the Embryonic Primate. *Cerebral Cortex*, 9(6):586–600.
- 917 Essen, D. C. V., Glasser, M. F., Dierker, D. L., and Harwell, J. (2011). Cortical  
918 Parcellations of the Macaque Monkey Analyzed on Surface-Based Atlases.  
919 *Cerebral Cortex*, page bhr290.



- 920 Fiddes, I. T., Lodewijk, G. A., Mooring, M., Bosworth, C. M., Ewing, A. D.,  
921 Mantalas, G. L., Novak, A. M., van den Bout, A., Bishara, A., Rosenkrantz,  
922 J. L., Lorig-Roach, R., Field, A. R., Haeussler, M., Russo, L., Bhaduri, A.,  
923 Nowakowski, T. J., Pollen, A. A., Dougherty, M. L., Nuttle, X., Addor, M.-C.,  
924 Zwolinski, S., Katzman, S., Kriegstein, A., Eichler, E. E., Salama, S. R., Jacobs,  
925 F. M. J., and Haussler, D. (2018). Human-Specific NOTCH2nl Genes Affect  
926 Notch Signaling and Cortical Neurogenesis. *Cell*, 173(6):1356–1369.e22.
- 927 Fisher, J. and Henzinger, T. A. (2007). Executable cell biology. *Nature Biotechnol-*  
928 *ogy*, 25(11):1239–1249.
- 929 Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner,  
930 C., Sykes, A., Wong, F. K., Peters, J., Guhr, E., Klemroth, S., Prufer, K., Kelso,  
931 J., Naumann, R., Nusslein, I., Dahl, A., Lachmann, R., Paabo, S., and Huttner,  
932 W. B. (2015). Human-specific gene ARHGAP11b promotes basal progenitor  
933 amplification and neocortex expansion. *Science*.
- 934 Florio, M., Namba, T., Pääbo, S., Hiller, M., and Huttner, W. B. (2016). A single  
935 splice site mutation in human-specific ARHGAP11b causes basal progenitor  
936 amplification. *Science Advances*, 2(12):e1601941.
- 937 Franco, S. J., Gil-Sanz, C., Martinez-Garay, I., Espinosa, A., Harkins-Perry, S. R.,  
938 Ramos, C., and Müller, U. (2012). Fate-restricted neural progenitors in the  
939 mammalian cerebral cortex. *Science (New York, N.Y.)*, 337(6095):746–749.
- 940 Franco, S. J. and Müller, U. (2013). Shaping Our Minds: Stem and Progenitor Cell  
941 Diversity in the Mammalian Neocortex. *Neuron*, 77(1):19–34.

- 942 Gao, P., Postiglione, M. P., Krieger, T. G., Hernandez, L., Wang, C., Han, Z.,  
943 Streicher, C., Papusheva, E., Insolera, R., Chugh, K., Kodish, O., Huang, K.,  
944 Simons, B. D., Luo, L., Hippenmeyer, S., and Shi, S.-H. (2014). Deterministic  
945 Progenitor Behavior and Unitary Production of Neurons in the Neocortex. *Cell*,  
946 159(4):775–788.
- 947 Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., Van  
948 Den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann,  
949 S. N., and others (2008). An intrinsic mechanism of corticogenesis from embry-  
950 onic stem cells. *Nature*, 455(7211):351–357.
- 951 Glass, L. and Kauffman, S. A. (1973). The logical analysis of continuous, non-  
952 linear biochemical control networks. *Journal of Theoretical Biology*, 39(1):103–  
953 129.
- 954 Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H., and Macklis,  
955 J. D. (2013). Molecular logic of neocortical projection neuron specification,  
956 development and diversity. *Nature Reviews Neuroscience*, 14(11):755–769.
- 957 Guo, C., Eckler, M. J., McKenna, W. L., McKinsey, G. L., Rubenstein, J. L. R.,  
958 and Chen, B. (2013). Fezf2 Expression Identifies a Multipotent Progenitor for  
959 Neocortical Projection Neurons, Astrocytes, and Oligodendrocytes. *Neuron*,  
960 80(5):1167–1174.
- 961 Götz, M. and Huttner, W. B. (2005). The cell biology of neurogenesis. *Nature*  
962 *Reviews Molecular Cell Biology*, 6(10):777–788.
- 963 Hartfuss, E., Galli, R., Heins, N., and Götz, M. (2001). Characterization of CNS  
964 precursor subtypes and radial glia. *Developmental Biology*, 229(1):15–30.

- 965 Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999). From  
966 molecular to modular cell biology. *Nature*, 402:C47–C52.
- 967 Haubensak, W., Attardo, A., Denk, W., and Huttner, W. B. (2004). Neurons arise  
968 in the basal neuroepithelium of the early mammalian telencephalon: A major  
969 site of neurogenesis. *Proceedings of the National Academy of Sciences of the*  
970 *United States of America*, 101(9):3196–3201.
- 971 Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L.,  
972 Miller, J. A., van de Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L.,  
973 Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena,  
974 P. M., Chakravarty, M. M., Chapin, M., Chong, J., Dalley, R. A., Daly, B. D.,  
975 Dang, C., Datta, S., Dee, N., Dolbeare, T. A., Faber, V., Feng, D., Fowler,  
976 D. R., Goldy, J., Gregor, B. W., Haradon, Z., Haynor, D. R., Hohmann, J. G.,  
977 Horvath, S., Howard, R. E., Jeromin, A., Jochim, J. M., Kinnunen, M., Lau, C.,  
978 Lazarz, E. T., Lee, C., Lemon, T. A., Li, L., Li, Y., Morris, J. A., Overly, C. C.,  
979 Parker, P. D., Parry, S. E., Reding, M., Royall, J. J., Schulkin, J., Sequeira, P. A.,  
980 Slaughterbeck, C. R., Smith, S. C., Sodt, A. J., Sunkin, S. M., Swanson, B. E.,  
981 Vawter, M. P., Williams, D., Wohnoutka, P., Zielke, H. R., Geschwind, D. H.,  
982 Hof, P. R., Smith, S. M., Koch, C., Grant, S. G. N., and Jones, A. R. (2012).  
983 An anatomically comprehensive atlas of the adult human brain transcriptome.  
984 *Nature*, 489(7416):391–399.
- 985 Haydar, T. F., Ang, E., and Rakic, P. (2003). Mitotic spindle rotation and mode  
986 of cell division in the developing telencephalon. *Proceedings of the National*  
987 *Academy of Sciences of the United States of America*, 100(5):2890–2895.

- 988 He, J., Zhang, G., Almeida, A. D., Cayouette, M., Simons, B. D., and Harris, W. A.  
989 (2012). How variable clones build an invariant retina. *Neuron*, 75(5):786–798.
- 990 Heins, N., Malatesta, P., Cecconi, F., Nakafuku, M., Tucker, K. L., Hack, M. A.,  
991 Chapouton, P., Barde, Y.-A., and Götz, M. (2002). Glial cells generate neurons:  
992 the role of the transcription factor Pax6. *Nature Neuroscience*, 5(4):308–315.
- 993 Huang, S., Guo, Y.-P., May, G., and Enver, T. (2007). Bifurcation dynamics  
994 in lineage-commitment in bipotent progenitor cells. *Developmental Biology*,  
995 305(2):695–713.
- 996 Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of  
997 proteins. *Journal of Molecular Biology*, 3(3):318–356.
- 998 Johnson, M. B., Kawasaki, Y. I., Mason, C. E., Krsnik, , Coppola, G., Bogdanović,  
999 D., Geschwind, D. H., Mane, S. M., State, M. W., and Šestan, N. (2009).  
1000 Functional and evolutionary insights into human brain development through  
1001 global transcriptome analysis. *Neuron*, 62(4):494–509.
- 1002 Kaplan, E. S., Ramos-Laguna, K. A., Mihalas, A. B., Daza, R. A. M., and Hevner,  
1003 R. F. (2017). Neocortical Sox9+ radial glia generate glutamatergic neurons for  
1004 all layers, but lack discernible evidence of early laminar fate restriction. *Neural*  
1005 *Development*, 12.
- 1006 Kauffman, S. A. and Kauffman, S. (1993). *The Origins of Order: Self-organization*  
1007 *and Selection in Evolution*. Oxford University Press.
- 1008 Kowalczyk, T., Pontious, A., Englund, C., Daza, R. A. M., Bedogni, F., Hodge,  
1009 R., Attardo, A., Bell, C., Huttner, W. B., and Hevner, R. F. (2009). Interme-

- 1010 diate Neuronal Progenitors (Basal Progenitors) Produce Pyramidal-Projection  
1011 Neurons for All Layers of Cerebral Cortex. *Cerebral Cortex*, 19(10):2439–2450.
- 1012 Krubitzer, L. and Kaas, J. (2005). The evolution of the neocortex in mammals:  
1013 how is phenotypic diversity generated? *Current Opinion in Neurobiology*,  
1014 15(4):444–453.
- 1015 Lange, C., Huttner, W. B., and Calegari, F. (2009). Cdk4/CyclinD1 Overexpression  
1016 in Neural Stem Cells Shortens G1, Delays Neurogenesis, and Promotes the  
1017 Generation and Expansion of Basal Progenitors. *Cell Stem Cell*, 5(3):320–331.
- 1018 Malatesta, P., Hack, M. A., Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff,  
1019 F., and Götz, M. (2003). Neuronal or glial progeny: regional differences in  
1020 radial glia fate. *Neuron*, 37(5):751–764.
- 1021 Mitchell, C. and Silver, D. L. (2018). Enhancing our brains: Genomic mechanisms  
1022 underlying cortical evolution. *Seminars in Cell & Developmental Biology*,  
1023 76:23–32.
- 1024 Miyata, T., Kawaguchi, A., Okano, H., and Ogawa, M. (2001). Asymmetric  
1025 Inheritance of Radial Glial Fibers by Cortical Neurons. *Neuron*, 31(5):727–741.
- 1026 Miyata, T., Kawaguchi, A., Saito, K., Kawano, M., Muto, T., and Ogawa, M.  
1027 (2004). Asymmetric production of surface-dividing and non-surface-dividing  
1028 cortical progenitor cells. *Development*, 131(13):3133–3145.
- 1029 Monuki, E. S. and Walsh, C. A. (2001). Mechanisms of cerebral cortical patterning  
1030 in mice and humans. *Nature Neuroscience*, 4:1199–1206.

1031 Ng, L., Bernard, A., Lau, C., Overly, C. C., Dong, H.-W., Kuan, C., Pathak, S.,  
1032 Sunkin, S. M., Dang, C., Bohland, J. W., and others (2009). An anatomic gene  
1033 expression atlas of the adult mouse brain. *Nature neuroscience*, 12(3):356–362.

1034 Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and  
1035 Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophecto-  
1036 derm differentiation. *Cell*, 123(5):917–929.

1037 Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S., and Kriegstein,  
1038 A. R. (2001). Neurons derived from radial glial cells establish radial units in  
1039 neocortex. *Nature*, 409(6821):714–720.

1040 Noctor, S. C., Flint, A. C., Weissman, T. A., Wong, W. S., Clinton, B. K., and  
1041 Kriegstein, A. R. (2002). Dividing precursor cells of the embryonic cortical  
1042 ventricular zone have morphological and molecular characteristics of radial  
1043 glia. *The Journal of Neuroscience: The Official Journal of the Society for*  
1044 *Neuroscience*, 22(8):3161–3173.

1045 Noctor, S. C., Martínez-Cerdeño, V., Ivic, L., and Kriegstein, A. R. (2004). Cortical  
1046 neurons arise in symmetric and asymmetric division zones and migrate through  
1047 specific phases. *Nature neuroscience*, 7(2):136–144.

1048 Noctor, S. C., Martínez-Cerdeño, V., and Kriegstein, A. R. (2008). Distinct  
1049 behaviors of neural stem and progenitor cells underlie cortical neurogenesis. *The*  
1050 *Journal of Comparative Neurology*, 508(1):28–44.

1051 Olariu, V., Coca, D., Billings, S. A., Tonge, P., Gokhale, P., Andrews, P. W., and  
1052 Kadirkamanathan, V. (2009). Modified variational Bayes EM estimation of  
1053 hidden Markov tree model of cell lineages. *Bioinformatics*, 25(21):2824–2830.

- 1054 O’Leary, D. D. M., Chou, S.-J., and Sahara, S. (2007). Area Patterning of the  
1055 Mammalian Cortex. *Neuron*, 56(2):252–269.
- 1056 Pfeiffer, M., Betizeau, M., Waltispurger, J., Pfister, S. S., Douglas, R. J., Kennedy,  
1057 H., and Dehay, C. (2016). Unsupervised lineage-based characterization of  
1058 primate precursors reveals high proliferative and morphological diversity in the  
1059 OSVZ. *The Journal of Comparative Neurology*, 524(3):535–563.
- 1060 Pilaz, L.-J., Patti, D., Marcy, G., Ollier, E., Pfister, S., Douglas, R. J., Betizeau,  
1061 M., Gautier, E., Cortay, V., Doerflinger, N., Kennedy, H., and Dehay, C. (2009).  
1062 Forced G1-phase reduction alters mode of division, neuron number, and laminar  
1063 phenotype in the cerebral cortex. *Proceedings of the National Academy of  
1064 Sciences*, 106(51):21924–21929.
- 1065 Polleux, F., Dehay, C., and Kennedy, H. (1997a). The timetable of laminar  
1066 neurogenesis contributes to the specification of cortical areas in mouse isocortex.  
1067 *The Journal of Comparative Neurology*, 385(1):95–116.
- 1068 Polleux, F., Dehay, C., Moraillon, B., and Kennedy, H. (1997b). Regulation of  
1069 Neuroblast Cell-Cycle Kinetics Plays a Crucial Role in the Generation of Unique  
1070 Features of Neocortical Areas. *The Journal of Neuroscience*, 17(20):7763–7783.
- 1071 Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental  
1072 biology. *Nature Reviews Neuroscience*, 10(10):724–735.
- 1073 Rakic, P., Ayoub, A. E., Breunig, J. J., and Dominguez, M. H. (2009). Decision by  
1074 division: making cortical maps. *Trends in Neurosciences*, 32(5):291–301.
- 1075 Rash, B. G. and Grove, E. A. (2006). Area and layer patterning in the developing  
1076 cerebral cortex. *Current Opinion in Neurobiology*, 16(1):25–34.

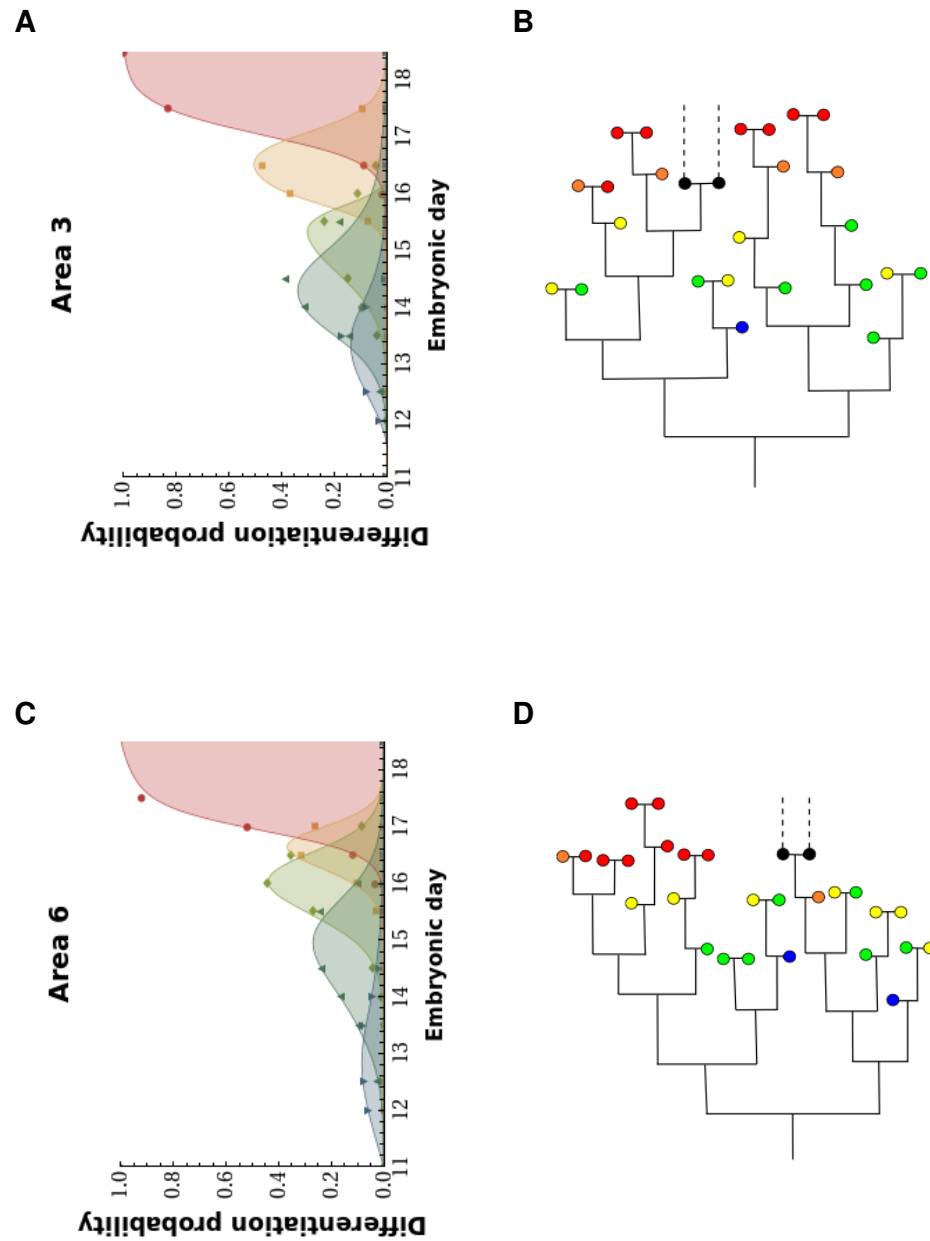
- 1077 Shen, Q., Wang, Y., Dimos, J. T., Fasano, C. A., Phoenix, T. N., Lemischka, I. R.,  
1078 Ivanova, N. B., Stifani, S., Morrissey, E. E., and Temple, S. (2006). The timing of  
1079 cortical neurogenesis is encoded within lineages of individual progenitor cells.  
1080 *Nature neuroscience*, 9(6):743–751.
- 1081 Striedter, G. F. (2005). *Principles of Brain Evolution*. Sinauer, Sunderland, MA.
- 1082 Sur, M. and Rubenstein, J. L. R. (2005). Patterning and Plasticity of the Cerebral  
1083 Cortex. *Science*, 310(5749):805–810.
- 1084 Suzuki, I. K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Her-  
1085 poel, A., Lambert, N., Cheron, J., Polleux, F., Detours, V., and Vanderhaeghen,  
1086 P. (2018). Human-Specific NOTCH2nl Genes Expand Cortical Neurogenesis  
1087 through Delta/Notch Regulation. *Cell*, 173(6):1370–1384.e16.
- 1088 Telley, L., Govindan, S., Prados, J., Stevant, I., Nef, S., Dermitzakis, E., Dayer, A.,  
1089 and Jabaudon, D. (2016). Sequential transcriptional waves direct the differentia-  
1090 tion of newborn neurons in the mouse neocortex. *Science*, 351(6280):1443–1446.
- 1091 Vasistha, N. A., García-Moreno, F., Arora, S., Cheung, A. F. P., Arnold, S. J.,  
1092 Robertson, E. J., and Molnár, Z. (2015). Cortical and Clonal Contribution of  
1093 Tbr2 Expressing Progenitors in the Developing Mouse Brain. *Cerebral Cortex*  
1094 (*New York, N.Y.: 1991*), 25(10):3290–3302.
- 1095 von Luxburg, U. (2007). A Tutorial on Spectral Clustering. *arXiv:0711.0189 [cs]*.  
1096 arXiv: 0711.0189.
- 1097 Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L.,  
1098 Pan, N., Xu, X., Tang, F., Zhang, J., Qiao, J., and Wang, X. (2018). A single-cell



- 1099 RNA-seq survey of the developmental landscape of the human prefrontal cortex.  
1100 *Nature*, 555(7697):524–528.
- 1101 Zubler, F. and Douglas, R. (2009). A framework for modeling the growth and  
1102 development of neurons and networks. *Frontiers in Computational Neuroscience*,  
1103 3:25.
- 1104 Zubler, F., Hauri, A., Pfister, S., Bauer, R., Anderson, J. C., Whatley, A. M., and  
1105 Douglas, R. J. (2013). Simulating Cortical Development as a Self Constructing  
1106 Process: A Novel Multi-Scale Approach Combining Molecular and Physical  
1107 Aspects. *PLoS Comput Biol*, 9(8):e1003173.
- 1108 Šestan, N., Rakic, P., and Donoghue, M. J. (2001). Independent parcellation of  
1109 the embryonic visual cortex and thalamus revealed by combinatorial Eph/ephrin  
1110 gene expression. *Current Biology*, 11(1):39–43.

## 1111 **8. Figures**

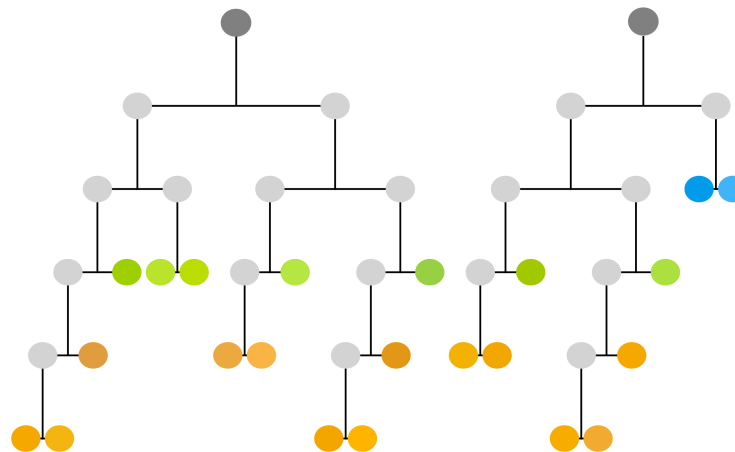
**Figure 1**



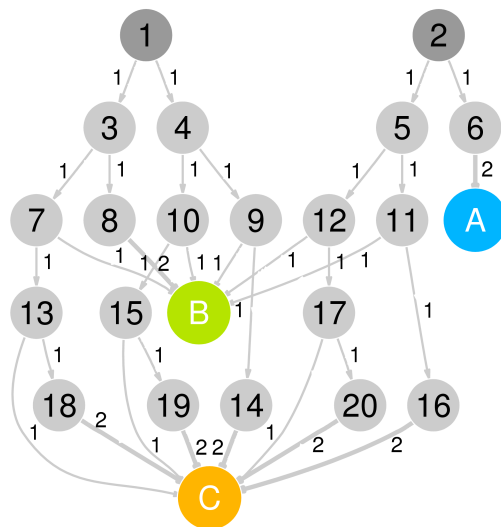
**Figure 1. Probabilistic generation of lineage trees.** Lineage trees are generated by sampling from the experimentally determined probability distribution (re-analysed from data of Polleux et al. (Polleux et al., 1997a)). **(A,C)** Probability distributions for area 3 and 6. Points, experimental data; lines, fits to data. **(B,D)** Example of sampled lineage trees. Trees layed out to correspond with the time axis of the experimental data. Black: precursor cell; blue: layer 6b; green: layer 6a; yellow: layer 5; orange: layer 4; red: layer 2-3; dashed lines, proliferation of glial precursor cells (not modeled).

**Figure 2**

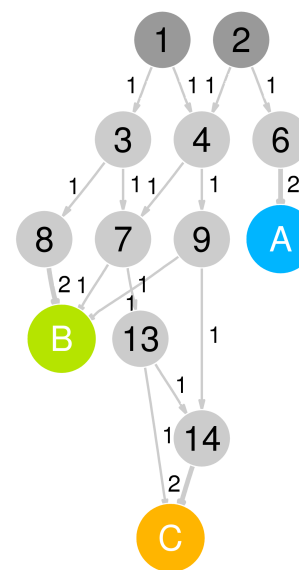
**A**



**B**



**C**

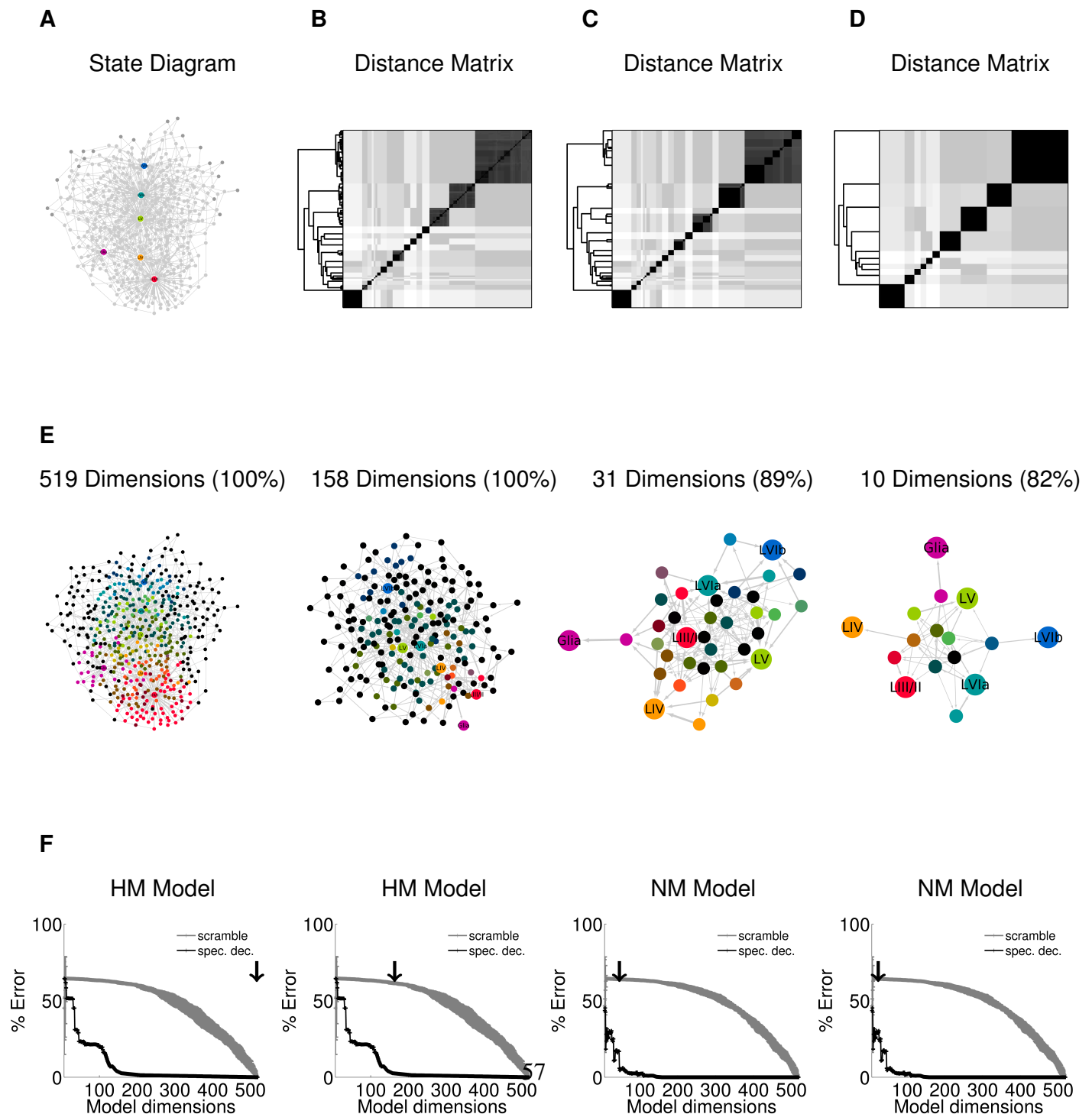


55

●  $\langle f_A=?, f_B=?, f_C=? \rangle$  ●  $\langle f_A=?, f_B=?, f_C=? \rangle$  ●  $\langle f_A=1, f_B=0, f_C=0 \rangle$  ●  $\langle f_A=0, f_B=1, f_C=0 \rangle$  ●  $\langle f_A=0, f_B=0, f_C=1 \rangle$

**Figure 2. Cell Lineage Trees and their corresponding State Diagram.** (A) Illustrative example of two cell lineage trees. Each node corresponds to a cell, and connecting edges to cell divisions. Two progenitor cells (dark gray) divide to form various hidden proliferative cells (light gray) and thereby give rise to 22 observable, terminally differentiated cells. Colors represent vectors of observed features  $\langle f_A, f_B, f_C \rangle$ . (B) State Diagram describes how the various cell states in lineage trees of A) are related. The hidden states are numbered in correspondence with each hidden cell in the lineages. Colored cells in the lineages have the same phenotypic features and so are represented by only a single state here. Edges between nodes indicate the transition probabilities  $p_{ij}$  from states  $i$  to  $j$  (the probabilities account for 2 offsprings per division). (C) Reduced State Diagram obtained by combining the redundant hidden states of B).

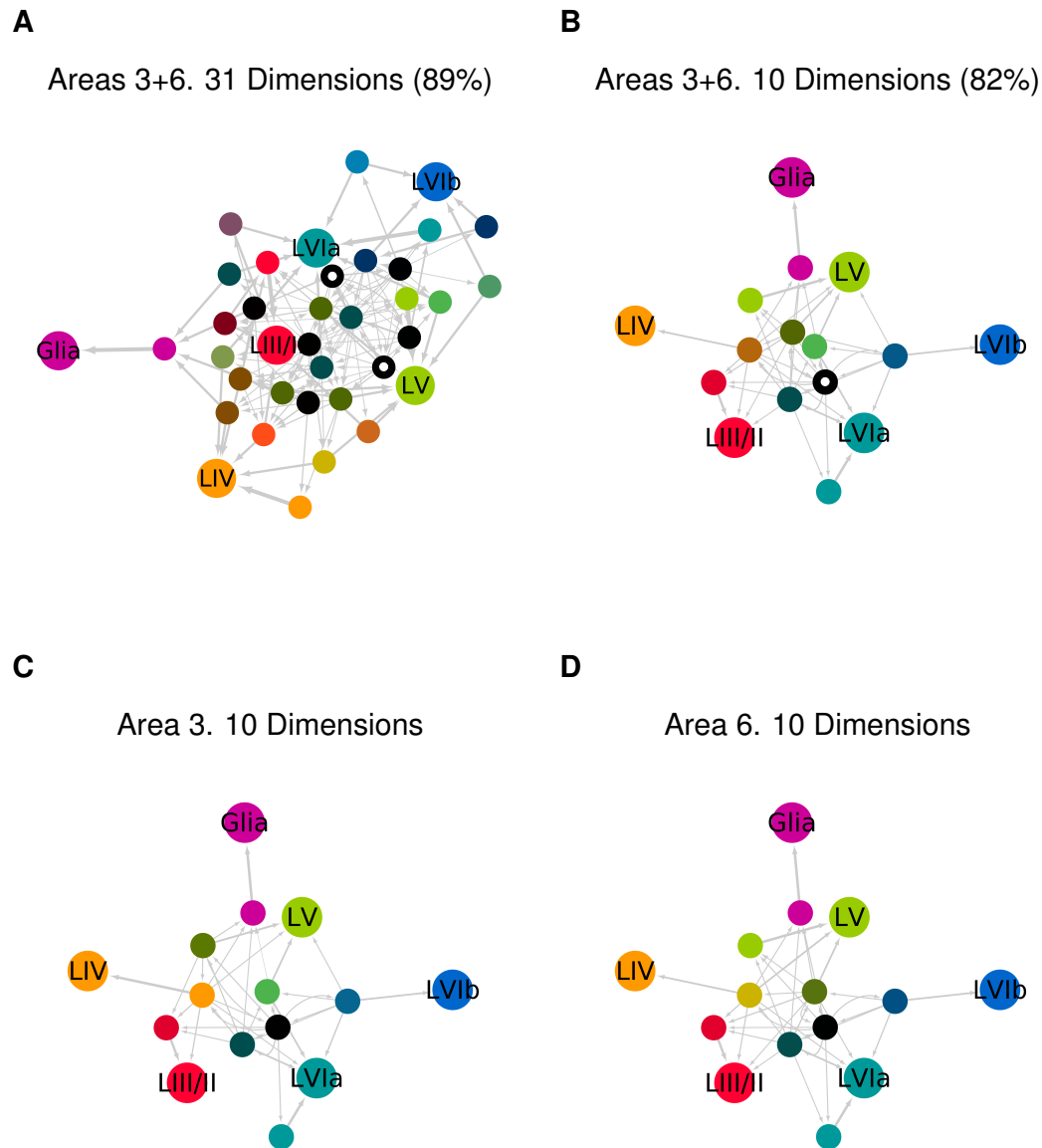
**Figure 3**



**Figure 3. State Diagram of cortical area 3 and 6.** (A) State diagram of cortical lineages in area 3 and 6 combined. Nodes represent cell states, arrows state transition probabilities. Cell states are labeled: blue: layer 6b; green: layer 6a; yellow: layer 5; orange: layer 4; red: layer 2-3; glia: pink, unknown; gray. Initial states are depicted as dark gray. (B-D) State clustergrams of computed distance between every state pair with dimensions  $D = 519$ ,  $D = 158$ ,  $D = 31$ , and  $D = 10$  (percentage of data represented in parenthesis). Dendrograms indicate hierarchical binary linkage of states. (E) Spectral label propagation on models, where each nodes is colored according to the estimated feature distribution. (F) Model error as percentage of the correct final cell states distribution for spectral clustering (black) versus random model (gray, standard deviations on 100 trials). HM, Homogeneous Markov model; NM, Non-Homogenous Markov Model. Black arrow indicates dimensionality of model.

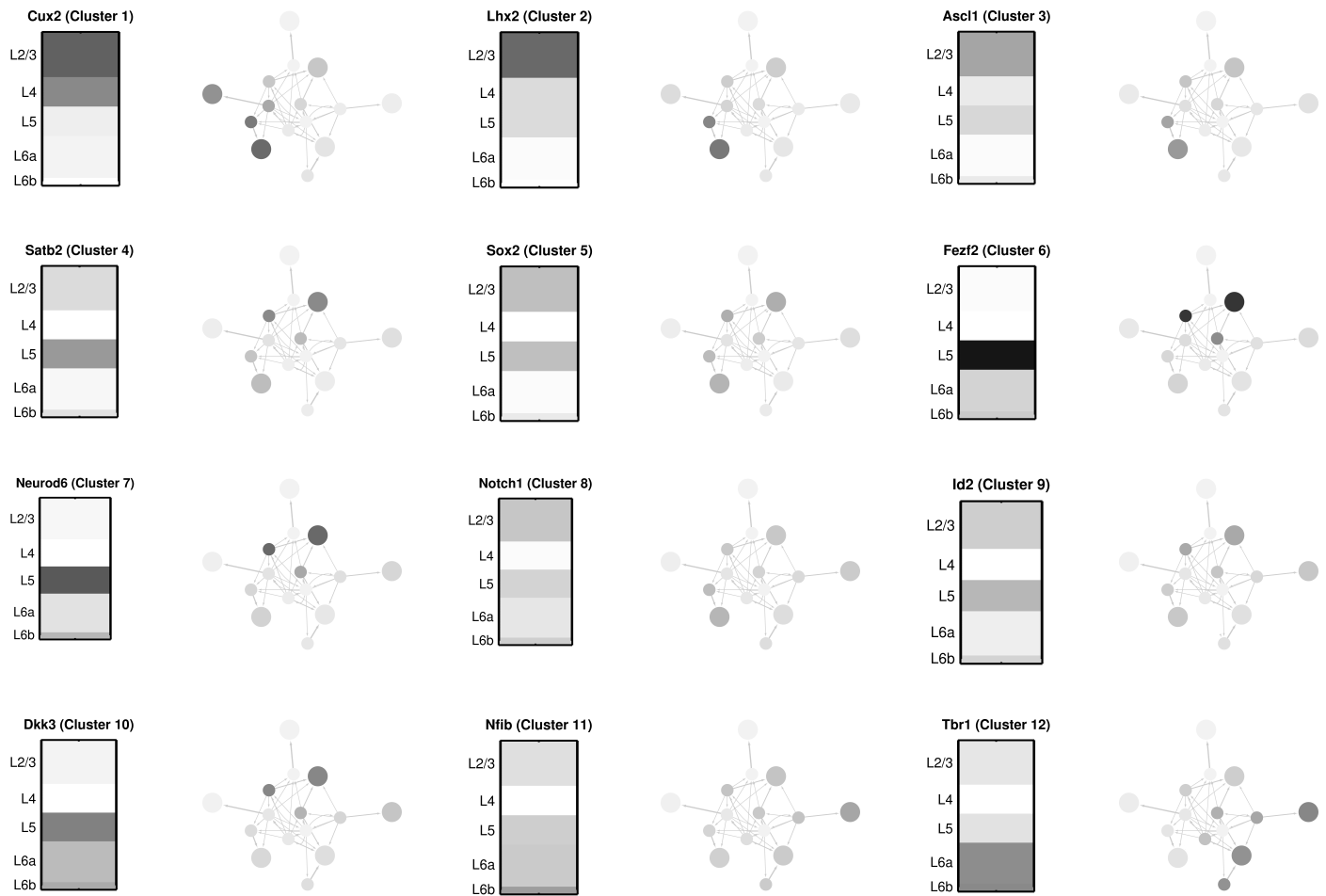


Figure 4



**Figure 4. State Diagram details.** (A-B) State Diagrams describing the combined lineages of areas 3 and 6. These 31 and 10 dimensional diagrams are enlarged from Figure 3. The initial precursor population(s) in these two cases are marked by centered white dots. The 31 dimensional SD declares a small second precursor population, whereas the 10 dimensional case collapses these two into a single initial population (with a small loss in ability to capture the experimental data). (C-D) Comparison of the two reduced State Diagrams for areas 3 and 6 respectively. The subtle differences can be seen in the shades of the three green/ocre small nodes in the upper left quadrants of the networks. The differences in shade indicate slight differences in predisposition towards terminal fates. (Networks enlarged from Suppl.Figures S6 and S7).

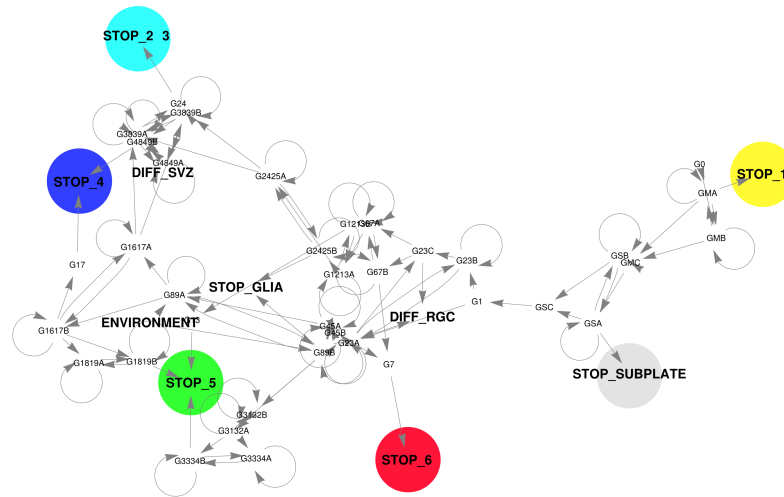
**Figure 5**



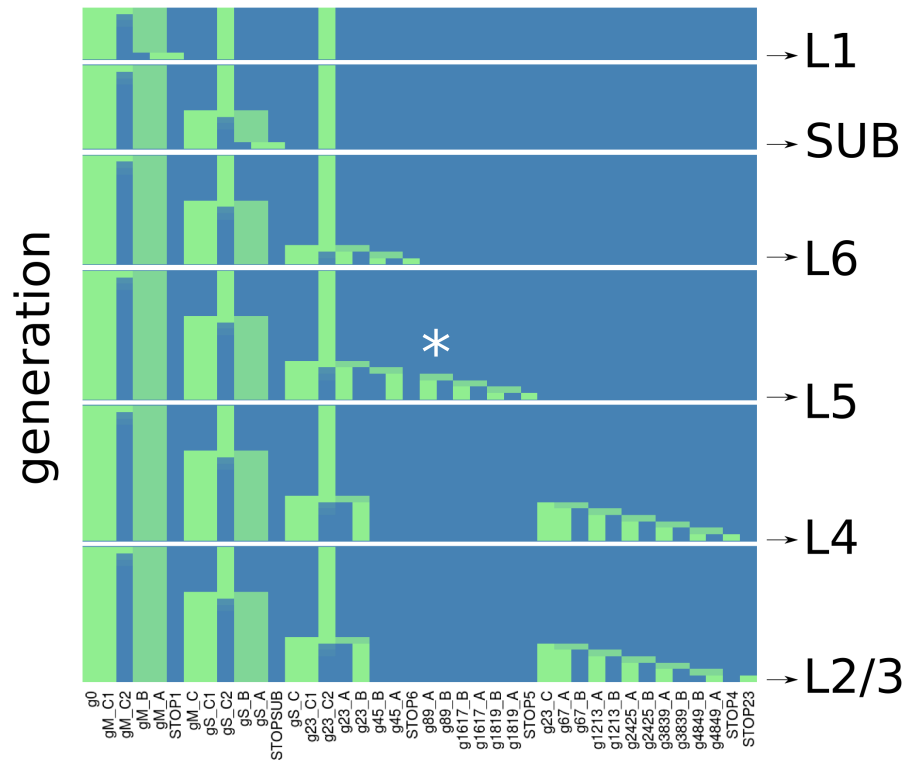
**Figure 5. Prediction of transcription factor expression across precursors** The expression patterns of 1751 transcription factors was measured in the adult mouse cortex by Belgard et al. (2011). We clustered these patterns into 12 groups according to similarity of their laminar distribution (see Table S1). The expression pattern of one representative factor from each group is shown in the 12 schematic cortical columns (grey value in proportion to observed expression). For each case, the adult expression pattern was assigned to the terminal states of the  $D = 10$  State Diagram (Figure 3). These values were propagated backwards into the SD as explained in the text. Grey shades of precursors indicate their predicted expression of that transcription factor. Thus, the 12 SDs together predict the profiles of expression of the 12 factors (and their groups) across all the cell states of neurogenesis as encoded by the State Diagram.

**Figure 6**

**A**

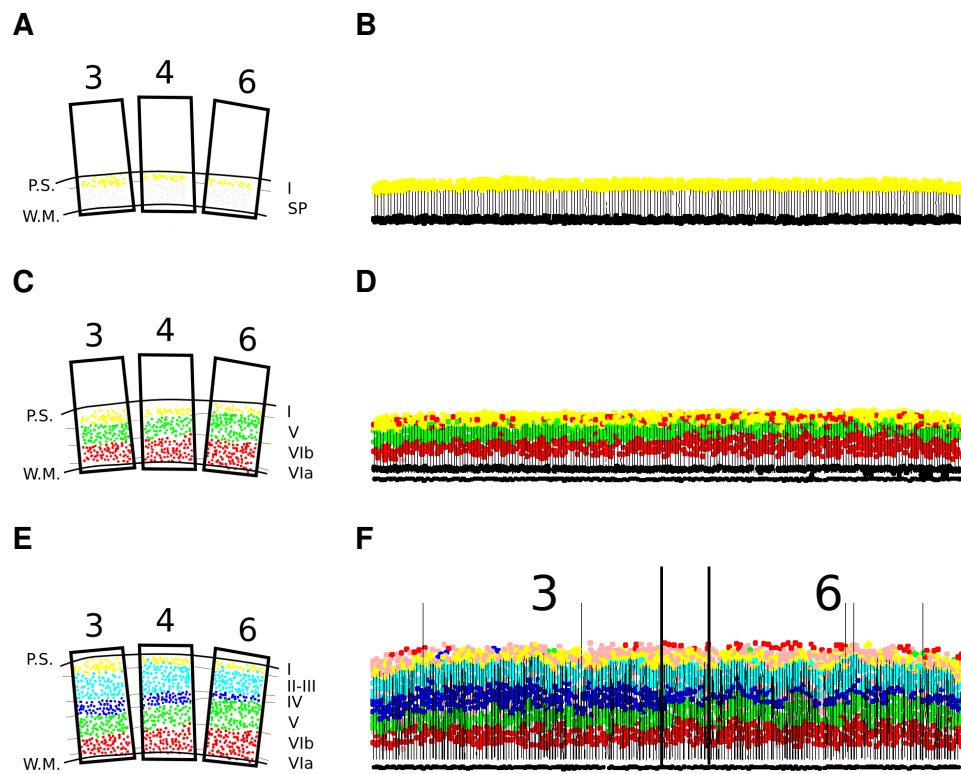


**B**



**Figure 6. GRN controlling simulated development of mouse cortex.** (A) Core Gene Regulatory Network controlling the production of marginal zone cells, and 5 different neuronal types of cortical area 3 and 6 in the mouse. Colored genes are expressed in neuron terminal states, and trigger differentiation. (B) Temporal expression pattern of core genes along lineages to 6 randomly selected cells of different type. Each panel shows the expression pattern of the initial precursor above, then patterns expressed by the next approximately 20 generations along lineage path, until terminal differentiating state is reached (below). Gene labels are shown beneath the lowest panel (L2/3). The expression patterns were measured immediately before mitosis, or at differentiation. At these times the genetic network reaches an attractor state. Expression levels range from 0 (blue) to 1 (green). Expression of gene ‘g89’, that biases neurogenesis towards either the area 3 or area 6 architectural phenotype, is indicated by white asterisk on path to layer 5 neuron.

**Figure 7**



**Figure 7. Simulation of cortical development.** (A-C-E) Schematic visualization of cortical area 3, 4, and 6 derived from 500  $\mu\text{m}$  paraffin sections counterstained with cresyl violet. Adapted from Polleux et al. (1997b). P.S., pial surface; W.M., white matter, SP, subplate. (B,D,F) Cx3D simulation of cortical development. For visualization, only a thin slice through the overall developing sheet is shown. (B) E11, with formation of marginal zone, subplate and radial glial cells; (D) E13, established infragranular layers; and (F) E16, established granular and supragranular layers, production of first glial cells. Area 3 and 6 boundaries marked by vertical black lines. There is a short transition zone between the 3 and 6 boundaries. Black: neuroepithelial cells; white/light gray: subplate cells; brown: intermediate precursors from subventricular zone; red: layer 6a and 6b; green: layer 5; blue: layer 4; cyan: layer 2/3; yellow: Marginal Zone or layer 1; pink: apoptotic cells; vertical lines, radial glia processes.

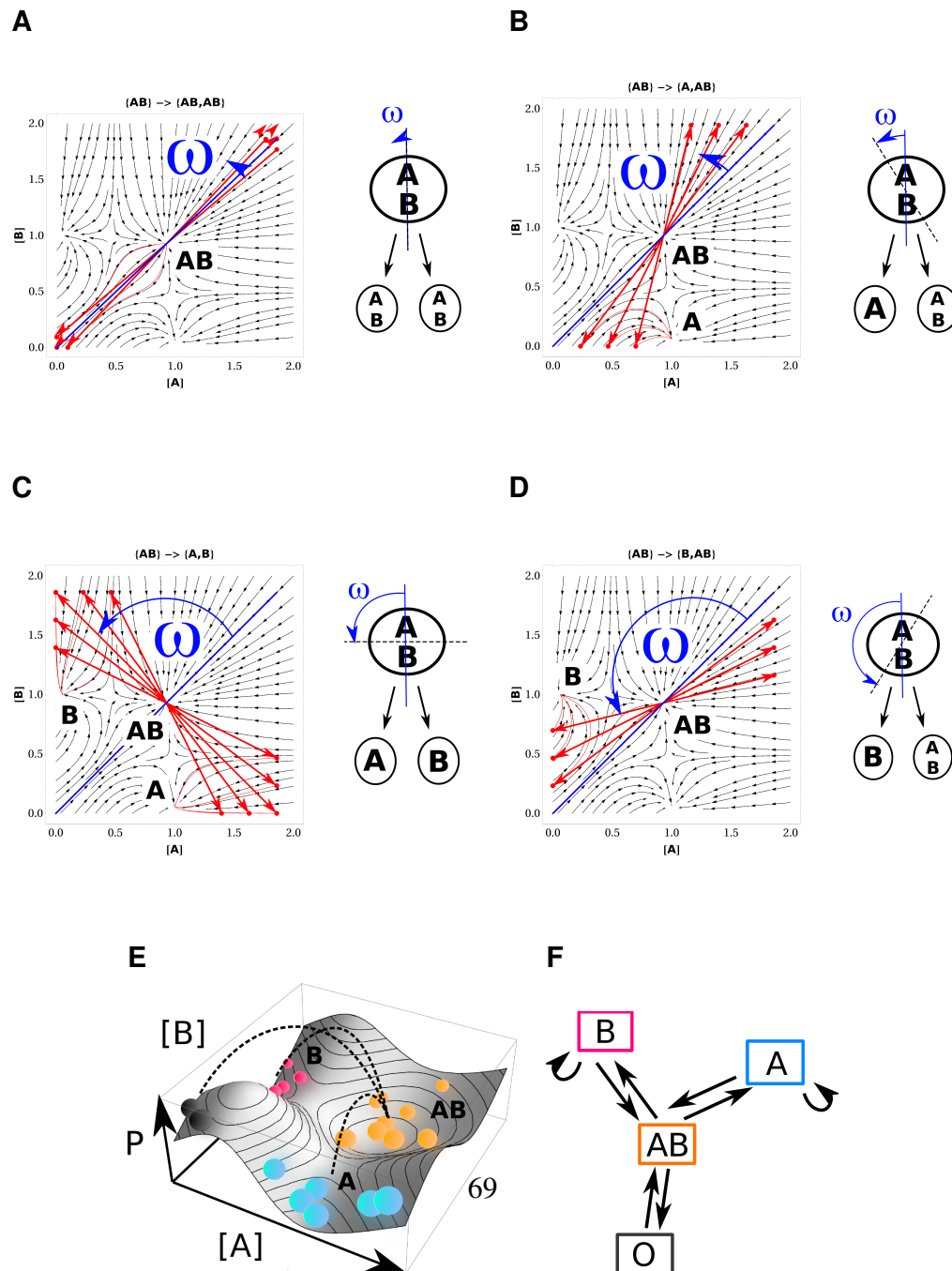


**Table 1**

<b>Layer</b>	<b>Area 6</b>		<b>Area 3</b>	
	<b>Experimental</b>	<b>Cx3D</b>	<b>Experimental</b>	<b>Cx3D</b>
<b>1</b>	0.9 ± 0.9	13.7 ± 0.0	1.2 ± 0.2	13.7 ± 0.0
<b>3-2</b>	27.1 ± 6.4	23.8 ± 3.8	28.4 ± 4.2	22.1 ± 3.5
<b>4</b>	12.0 ± 2.1	12.3 ± 2.7	19.7 ± 5.6	20.5 ± 3.1
<b>5</b>	27.0 ± 6.0	24.9 ± 3.1	18.6 ± 1.4	17.8 ± 2.3
<b>6</b>	32.9 ± 3.8	27.0 ± 4.1	33.5 ± 0.7	26.0 ± 4.5

**Table 1. Laminar distributions of differentiated cells.** Cells produced by simulations of GRN guided neurogenesis in areas 3 and 6. Quantification of simulated final neuronal production in each layer (before apoptosis) are compared with experimental data (Polleux et al., 1997a). Values are given in % with standard deviation. Experimental values were averaged and normalized to 100%.

**Figure 8**



**Figure 8. Genetic attractor landscape of a bistable switch with asymmetric cell division.**

Distributions of different division types as a function of division angle  $\omega$ . Different division patterns arise: **(A)**  $\{AB\} \rightarrow \{AB\}, \{AB\}$ ; **(B)**  $\{AB\} \rightarrow \{A\}, \{AB\}$ ; **(C)**  $\{AB\} \rightarrow \{A\}, \{B\}$ ; **(D)**  $\{AB\} \rightarrow \{A\}, \{AB\}$ . Red straight traces are simulated jumps at different angles, and red curvilinear trajectories show the time evolution after the jump. Blue lines indicate the  $\omega$  angle with respect to the internal distribution of proteins. **(E)** Schematic representation of an attractor landscape  $P$  as a function of the concentrations of two genes  $A$  and  $B$ , in absence of an input stimulus. The landscape is determined by the manner of interaction between the genes. Each point on landscape corresponds to a possible gene expression profile. Spheres correspond to cells in different attractor basins; dotted lines to possible state transitions. **(F)** State diagram of bistable switch. Transitions are possible only by influence of the expression of another gene (e.g. through input  $I$ , Figure S9), or asymmetric cell division.

## 1112 **9. Supporting Information: Tables**

## Table S1

<b>Cluster 1</b>	Barx2, Batf2, Bhlhe22, Cited4, Cux1, <b>Cux2</b> , Egr4, Emx2, Fgf2, Foxc1, Foxp3, Hmgn5, Hnf1a, Hsf4, Inhba, Konh4, Kcnh5, Klf2, Luc7l3, Maf, Mef2c, Mxk, Neurod1, Neurog3, Nkx3-1, Nog, Npnt, Nr2f1, Pou6f1, Pparg, Rbfox3, Rbms1, Rora, Rorb, Sox4, Tshz1, Wnt10a, Zfhx4, Zfp459, Zfpm1
<b>Cluster 2</b>	0610031J06Rik, 6030422M02Rik, Ablim2, Aes, Akap8l, Arid4a, Atrx, Bbx, Cacna1a, Camk2a, Camta2, Cc2d1a, Ccdc112, Chd2, Chd5, Cited2, Crtc1, Csdcd2, Dand5, Dapk3, Dbp, Dek, Dlg4, Dmrta2, Dnajc1, Edn1, Egr3, Ehmt2, ElI3, Emx1, Eng, Ercc2, Fosl2, Foxf2, Foxo3, Foxp1, Fzd1, Gcfc1, Gtf2f1, H1fx, H2afj, Hdac7, Hes5, Heyl, Hivep3, Ikzf4, Ing2, Irf7, Jdp2, Jund, Kcnh3, Kctd1, Khdrbs2, Kif5c, Klf13, <b>Lhx2</b> , Lmo4, Mapk11, Maz, Mbd3, Med25, Med29, Mli5, Mlt1, Mt3, Mtf2, Mxd4, Mybbp1a, Mzf1, Nfic, Nfix, Notch3, Nr2f6, Pbxip1, Pias4, Pim1, Pkn1, Poll, Polr2e, Polr2i, Ppargctb, Ppp3ca, Prox2, Ptov1, Ptrf, Rbck1, Recql5, Rere, Rfc5, Rrp8, Rsf1, Sap25, Scand1, Scrt1, Setbp1, Smad3, Smarca2, Smarcd3, Snapc4, Sox17, Sox18, Ssbp4, Ssrp1, Taf3, Tceal7, Tcf4, Thap3, Thap7, Tle3, Trerf1, Trim28, Ttf1, Usp2, Vgll4, Wfs1, Wnt4, Wnt9a, Zbtb46, Zfp316, Zfp329, Zfp444, Zfp462, Zfp523, Zfp575, Zfp579, Zfp628, Zfp771, Zfp821, Zfp827
<b>Cluster 3</b>	2310045N01Rik, Acd, Actl6b, Agap2, Agt, Ahdc1, Akt2, Ankrd49, Arid1a, Arid3b, Arid4b, <b>Ascl1</b> , Atf5, Atf6b, Atn1, Banf1, Bcl9l, Bmp7, Bptf, Brd2, Brd3, Brms1, Cand2, Cbfa2t3, Cck, Ccnt2, Cdk5r1, Cdk9, Cdkn1c, Cenpb, Chd4, Chd8, Cic, Cnot3, Crebbp, Crtc3, Ddb2, Ddit3, Ddx21, Ddx41, Deaf1, Dot1l, Drap1, Dvl1, Dyrk1b, Eli, Elof1, Erf, Esf1, Fbxl19, Fbxw7, Fiz1, Flywch1, Foxq1, Frzb, Fzd2, Gm9887, Golga4, Gsk3a, Gtf2ird2, H2afx, Hdac5, Hic2, Hras1, Ighmbp2, Impdh1, Ing1, Ing4, Ino80b, Irf2, Irf2bp1, Jhdm1d, Jmjd6, Kcnh2, Kdm5a, Klf16, Klf7, Ldb1, Lig1, Lmna, Lmo1, Lrp5, Lyl1, Maml3, Map3k10, Mcrs1, Med19, Mli1, Mli2, Mtap1s, Mtdh, Mxd3, Mypop, Naa15, Nat14, Ncor1, Ndufa13, Nedd8, Nfil3, Nfkbia, Npas4, Nr2e1, Nr4a1, Paf1, Pcbp4, Pde8b, Per1, Per3, Phc2, Phf12, Phip, Pkd2, Polr2j, Ppp1r12a, Preb, Prr13, Psen2, Psp1, Rad54l, Rai1, Rbpj, Rdbp, Rfx1, Ring1, Rnf10, Rnf20, Rnf31, Rtf1, Rxrb, S100a1, Safb2, Samd1, Sdpr, Sec14l2, Sertad1, Set, Sirt7, Sltm, Smarca4, Smg6, Snapc2, Snw1, Sox11, Sox12, Sox9, Spen, Srrm1, Srsf10, Tada3, Taf10, Taok2, Tcea2, Tnrc18, Traf2, Trapp, Ubtfl, Upf1, Usp16, Usp21, Vps72, Wbp7, Xpa, Ybx1, Yy1, Zbed3, Zbtb17, Zbtb7a, Zbtb8a, Zlat, Zfhx2, Zfp148, Zfp213, Zfp219, Zfp414, Zfp513, Zfp524, Zfp580, Zfp641, Zfp768, Zfp777, Zfp787, Zfp825, Zfp865, Zgpl1, Zgpat, Zkscan17, Zmiz2
<b>Cluster 4</b>	0610010F05Rik, 1700048O20Rik, 2210018M11Rik, 2310047B19Rik, Ablim3, Acvr1b, Akap8, Akt1, Apbb2, Aptx, Arid1b, Arid5b, Arnt2, Arntl, Arrb1, Ash1l, Asxl1, Atmin, Atp6v0a1, Bach2, Bclaf1, Bdp1, Becn1, Brca2, Btaf1, C230052I12Rik, Calcooco1, Calr, Camk1d, Camta1, Carm1, Casp8ap2, Cbfa2t2, Cbx7, Cdk13, Cdkn1b, Cebpg, Cep290, Ciao1, Cnot4, Cnot7, Commd6, Coq9, Cry2, Csnk2a1, Csrnp2, Ctbp1, Dab2ip, Ddx52, Dmtf1, Dnajb5, Dnttip1, Dnttip2, Dpf1, Dpf2, E2f3, Ecsit, Eid2, Eif4g3, Ern1, Esrra, Fancm, Fbxw11, Fmn1, Fosb, Foxk2, Fzd4, Fzd6, Gatad1, Gatad2a, Gm20517, Grif1, Gsk3b, Gtf2a2, Gtf2f2, Gtf2h1, Gtf2h4, Gtf2h5, Gtf3c4, H2afz, Hcfc1, Hdac3, Hdac8, Hexim1, Hif1an, Hinfp, Hist3h2a, Hlf, Hmg20a, Hmga1, Hmgn3, Hnrnpd, Hnrnpu, Hnrpdll, Homez, Iws1, Jarid2, Jrk, Kat5, Kcnh7, Kdm1a, Kras, L3mbtl3, Leo1, Lrrfip1, Maged1, Map3k9, Mapre3, Mcm9, Mdm2, Med1, Med12l, Med13, Med15, Med18, Med27, Men1, Mrpl12, Msh3, Mtpn, Myh9, Ncoa1, Ncoa2, Nlk, Nom1, Npas2, Nr1d1, Nr1i3, Nrip1, Nsd1, Nufip1, Nusap1, Orc2, Paip1, Parp2, Paxip1, Pcgf3, Pcgf6, Pcid2, Pdcod4, Pdgbf, Pdpk1, Peo1, Per2, Pex14, Pgr, Phb2, Plk3r1, Plcb1, Polb, Poldip2, Poli, Polr1a, Polr3d, Polrmt, Pou3f3, Ppm1f, Ppp2r5b, Ppp2r5d, Prdm4, Prdx2, Prim2, Prkrir, Prmt6, Prmt7, Prpf19, Prpf6, Psma6, Psmc5, Psmnd10, Psmnd9, Ptges2, Pygo1, Rad1, Rad50, Rad51l3, Rbbp7, Rbm15, Rhoq, Rnf4, Rnf6, Rps6ka3, Rptor, S1pr1, Sap130, Sap30, <b>Satb2</b> , Scrt2, Setd3, Smc5, Smug1, Smyd2, Srap, Srxn1, Supv3l1, Tada2b, Taf11, Taf1b, Taf5l, Taf8, Tagln3, Taok1, Tbl1x, Tbp1l, Tceb1, Tceb3, Tceerg1, Tcf25, Tdg, Tgfb3, Tgfb3r, Thap4, Thrb, Tcam1, Tigd2, Tmem18, Tnfrsf11a, Top3a, Topors, Tox3, Trim37, Ube3a, Vegfa, Vldlr, Vps25, Wnt2b, Wwc1, Wwp2, Xrcc5, Yaf2, Ywhab, Ywhah, Zbtb25, Zbtb8b, Zbtb9, Zeb1, Zfp105, Zfp187, Zfp202, Zfp238, Zfp239, Zfp251, Zfp273, Zfp334, Zfp369, Zfp410, Zfp422, Zfp451, Zfp472, Zfp511, Zfp512, Zfp532, Zfp566, Zfp612, Zfp64, Zfp784, Zfp788, Zfp866, Zfp933, Zfp941, Zfp942, Zfp959, Zhx3, Zxdb
<b>Cluster 5</b>	1810035L17Rik, 2310004N24Rik, 2410016O06Rik, 2410022L05Rik, 2610301G19Rik, 4933421E11Rik, Abt1, Akna, Ankrd33b, Anp32a, Apbb1, Apex1, Ar, Arid2, Ascc1, Atf7ip, Atf7ip, Atp8b1, Atxn1, Atxn1l, Atxn7l3, Bag1, Bahd1, Baz1b, Baz2a, Bcl6b, Bcl9, Bcor, Bmyc, Bod1l, Brf2, Brwd1, C80913, Camk4, Camsap3, Cbx1, Cby1, Ccar1, Ccnk, Ccnt1, Cdk12, Cdk5, Cdk8, Cebpz, Chd1, Chmp1a, Chrac1, Chtf8, Cobra1, Cramp1l, Creb1, Creb3, Cry1, Csd4, Csnk2a2, Ctcf, Ctnnd2, Cxxc1, Cxxc5, Daxx, Dbx2, Dcaf6, Ddx17, Ddx50, Ddx54, Ddx56, Dedd2, Dlx1, Dmap1, Dusp22, Dvl3, E2f4, E2f5, E430018J23Rik, Ecd, Egr1, Eif2c1, Elk1, Elk4, Eme1, Ep400, Epas1, Epc2, Ercc1, Ercc4, Ercc5, Fbxo18, Fhod1, Fli1, Fosl1, Foxg1, Foxj2, Foxo1, Fus, Gli2, Gli3, Glo1, Gm6563, Gmcl1, Gmeb1, Gmeb2, Gon4l, Gtf3a, Gtf3c2, H2afy2, Hdac11, Hdac4, Hdgf, Hdgrp2, Hipk1, Hira, Hist3h2ba, Hivep2, Hnrnp1, Hnrnpul1, Hr, Hsf1, Htatsf1, Hyal2, Ifit4, Igf1, Ikbkap, Ilf2, Impdh2, Ing5, Ino80, Jmy, Kat8, Kcnh1, Kdm2a, Kdm4b, Kdm5b, Keap1, Khsp9, Klf15, Klf5, Klf6, Klf9, L3mbtl2, Lcor, Lonp1, Maf1, Mafg, Mamld1, Mapk14, Max, Mcm5, Mcm7, Mcts2, Mecp2, Med13l, Med26, Med28, Med9, Mef2d, Meis3, Mier2, Mki1, Mnat1, Morf4l1, Mpg, Mphosph8, Mta1, Mta2, Mxi1, Myd88, Mysm1, Myst3, Nacc1, Narfl, Nbn, Ncl, Ncor2, Ndp, Neurod2, Nfat5, Nfe2l1, Nfrkb, Nipbl, Nolc1, Npas1, Nr2c2, Nrarp, Nucb1, Nup62, Obfc2b, Ogg1, Otud7a, Pa2g4, Patz1, Pbx2, Pcbp3, Pdcod11, Pds5b, Phb, Phf1, Phf5a, Pias1, Pkd1, Plagl2, Pogz, Pole3, Polg, Polr1c, Polr1d, Polr2c, Polr2f, Polr2l, Polr3h, Pot1b, Ppap2b, Ppard, Pprc1, Pqbp1, Prdm12, Prdm2, Prdx5, Prkcz, Prmt5, Prr12, Psmd4, Puf60, Pura, Purg, Rad54l2, Rai12, Rb1cc1, Rbbp4, Rbm39, Rcor1, Rcor2, Rfc1, Rfc2, Rfc4, Rfxkank, Rfxap, Rnf187, Rprdlb, Rps6ka4, Safb, Sap30bp, Sap30l, Sbnol, Senp2, Setd2, Sf1, Sfsawap, Ski, Smarcb1, Smarcc1, Smarcc2, Smarcd1, Smo, Snip1, Son, <b>Sox2</b> , Sox21, Sra1, Srrt, Ssbp3, Stat5b, Stk16, Strn3, Suds3, Supt5h, Swap70, Taf5, Taf6, Tceal5, Tef, Terf2, Tfip11, Thap11, Thoc1, Thrsp, Tinf2, Top1, Tox4, Traf7, Trim27, Trp53bp1, Tsc22d1, Tshz3, Tsn, Tspyl2, Ube2l3, Ubqln4, Upf2, Usf2, Vps36, Wdr5, Wdtdc1, Whsc1l1, Whsc2, Wnt7a, Wrnip1, Wwtr1, Xbp1, Xpc, Xrcc1, Ylpm1, Zbtb22, Zbtb3, Zbtb38, Zfand3, Zfp113, Zfp119a, Zfp160, Zfp174, Zfp180, Zfp235, Zfp263, Zfp28, Zfp286, Zfp319, Zfp498, Zfp553, Zfp574, Zfp592, Zfp61, Zfp629, Zfp653, Zfp668, Zfp672, Zfp687, Zfp689, Zfp746, Zfp809, Zfp81, Zfp810, Zfp867, Zfp954, Zfr, Zkscan14, Zkscan4, Zmat2, Znf1x, Zscan29, Zzz3
<b>Cluster 6</b>	Aff3, Ahr, Aifm2, Ankrd42, Arx, Bcl6, Bhlhe40, Bhlhe41, Bmp2, Ccnh, Ctbp2, Cxxc4, Dusp5, Elp4, Esrrg, Etv1, <b>Fezf2</b> , Gas6, Hat1, Hes1, Il4, Lmo3, Msh2, Nck1, Nkrf, Nr1d2, Nrip2, Obfc2a, Parp1, Phf6, Ppargc1a, Prdx3, Prkaa2, Ralgapa1, Reln, Rgmb, Rnf14, Sall2, Satb1, Shh, Sla2, Smad9, Snappc3, Sod2, Tfb1m, Tgfbfr1, Tox, Tox2, Trib2, Tsc22d3, Uchl5, Zc3h8, Zfp260, Zfp367, Zfp458, Zmat4
<b>Cluster 7</b>	1500003O03Rik, 2700050L05Rik, Aifm1, Arhgef11, Atf4, Blm, Brms1l, Btrc, Cand1, Cask, Cd38, Cdk7, Cops2, Cops5, Crib1a1, Crebl2, Crem, Cseal1, Csrnp3, Ddx1, Ddx3x, Dnaj3, Dpy30, Dr1, E2f6, Eif4g2, Eme2, Ets2, Fam120b, Fbxo11, Fgfr3, Fzd9, Glyrl, Gm14296, Gm14326, Gpbbp1, Grm5, Gtf2b, Gzf1, Has3, Hey1, Hif1a, Hmbox1, Hmx5, Hsp8a, Igbp1, Ikbkg, Ilf6, Insig2, Klf12, Lassa4, Lbh, Lig4, Lonp2, Lpin2, Lrrppr, Mafb, Map3k13, Mcts1, Med14, Med21, Med30, Med31, Mix, Msh6, Mterfd3, Mtor, Ncoa7, Ndn12, <b>Neurod6</b> , Nfyb, Nif3l1, Nr3c2, Phf17, Pid1, Pole4, Polr1b, Polr3a, Polr3f, Polr3k, Pou3f4, Prkaa1, Psmc3ip, Ptch1, Ptpkr, Rabgef1, Rad23b, Rbfox2, Rpa1, Rpap2, Rqcd1, Rrn3, Setd7, Slc30a9, Sos1, Sfrbp1, Ss18l1, Strap, Taf2, Taf9, Tax1bp1, Tceal1, Terf2ip, Tmf1, Traf3, Trim32, Txlng, Uba3, Ube2b, Ube2n, Ubqln1, Wwp1, Yeats4, Zbtb10, Zbtb16, Zfp248, Zfp27, Zfp35, Zfp426, Zfp599, Zfp647, Zfp655, Zfp7, Zfp703, Zfp759, Zfp786, Zfp9, Zfp940, Zfp943, Zkscan1
<b>Cluster 8</b>	2210012G02Rik, 2700060E02Rik, 9130019O22Rik, A43003K04Rik, Abl1, Abim1, Adnp, Alyref, Alyref2, Arnt, Atf1, Atxn7, AW146020, Bmp6, Brd7, Btg2, C130039O16Rik, Capn3, Cbfb, Cbx4, Cdc5l, Cdk5rap3, Cebpba, Cebpba, Cenpt, Chd3, Chtf18, Clpb, Clu, Cnot6, Commd7, Crebzf, Ctdsp1, Ctnnd1, Cyld, Dap, Ddx39b, Dicer1, Dnajb6, Dnmt3a, Dvl2, Edf1, Eepd1, Eglin1, Elf1, Ewsr1, Foxk1, Foxo4, Foxp4, Fzd3, Gm10093, H2afv, H2afy, Hip1, Hipk2, Hist1h1c, Hist2h2aa1, Hmgb1, Hopx, Hp1bp3, Id1, Ifnar2, Ifit57, Ilk, Irf9, Jun, Junb, Kdm5c, Kdm6b, Limd1, Malt1, Maml2, Map2k1, Mapk13, Mapk8ip1, Mcf2l, Mli3, Mmp114, Mnt, Mmp5, Mst4, Mstl1, Nabd2, Naca, Nfe2l2, Nfkb2, Nod1, <b>Notch1</b> , Nras, Rnf1, Nrc3, Nucb2, Pask, Pbrm1, Pcnal, Pncb2, Pdnf5, Pfn1, Pknox1, Plag1, Plogk, Pola2, Polm, Ppp1r10, Prkch, Rbak, Rbpjl, Rela, Rgs14, Ripk1, Rpa2, Rps3, Rps6ka1, Ruvbl1, Sbnol, Scap, Scfm1, Sertad2, Setd1b, Sfqp, Sfrp1, Sin3b, Smad4, Sorbs3, Sox15, Sp9, Srf, Stat3, Stat5a, Sub1, Taf9, Tceal1, Tceb2, Tcf15, Tesc, Tfcop21, Tgfb1, Tgfb2, Thra, Thrap3, Tigd3, Trim11, Trps1, Tsc22d4, Tsnax, Ube2i, Ubp1, Usf1, Vhl, Vopp1, Xrcc6, Ywhaq, Zeb2, Zfp161, Zfp276, Zfp282, Zfp361l, Zfp40, Zfp41, Zfp438, Zfp473, Zfp521, Zfp536, Zfp560, Zfp652, Zfp710, Zfp772, Zfp811, Zhx2
<b>Cluster 9</b>	2610008E11Rik, Abtb2, Adar, Ad1, Af14, Arhgef2, Ascc2, Asf1a, Atf7, Atxn3, Axin1, Basp1, Bcl11a, Brd8, Brf1, Chaf1a, Cnbp, Ctif, Ctnnbip1, Dcp1a, Ddx5, Dedd, Dmd, Dnmt1, E2f1, Eapp, Eif2a, Ep300, Epc1, Fer, Fgf1, Fhl2, Flii, G3bp1, Gatad2b, Gm9833, Gpbbp1l1, Gtf3c1, H3f3b, Hacc1, Hbp1, Hes6, Hipk3, Hist1h2bc, Hist2h2be, <b>Id2</b> , Irak3, Irf8, Itch, Khdrbs1, Klf11, Klf3, Lass5, Lass6, Loxl3, Lrp6, Lrp8, Lrwd1, Mafk, Mapk1, Mbd2, Med24, Mms19, Mtf1, Ncoa6, Neo1, Nfatc3, Npas3, Nr3c1, Orc4, Orc6, Pcbp1, Pelil, Phf10, Phf2, Phf8, Ppm1a, Ppp1r8, Prkd1, Psen1, Pxmp3, Rb1, Rbl2, Rbm14, Rc3h2, Recql, Rev1, Rhoa, Rnf141, Rnf2, Ruvbl2, Ryr2, Sin3a, Smad1, Smad5, Smarca5, Snd1, Snrnp200, Sos2, Sp1, Sp4, Spin1, Srebf1, Srebf2, Supt6h, Suv420h1, Taf12, Taf4a, Tlap4, Tgfbpar1, Th1l, Thap2, Trak2, Trip4, Ttn1, Uhrf2, Usp22, Wasl, Xrn2, Zbtb5, Zland5, Zland6, Zfp108, Zfp110, Zfp119b, Zfp146, Zfp192, Zfp287, Zfp3, Zfp46, Zfp516, Zfp52, Zfp59, Zfp709, Zfp775, Zfp903, Zfp91, Zkscan18, Zxdc
<b>Cluster 10</b>	1810074P20Rik, 3110052M02Rik, A530054K11Rik, AA987161, Abcg1, Actr8, Adnp2, Aebp2, Akirin2, Aplp2, App, Ascc3, Atf2, Atf6, AW146154, Birc2, Bmp1ra1, Brdt, Bzw1, Carf, Cbx5, Ccpq1, Cdc73, Cenpc1, Cggbp1, Cih1a, Clpx, Cnot1, Cnot2, Cnot8, Commd1, Csrnp1, Ddb1, Ddx20, <b>Dkk3</b> , Eaf1, Ednrb, Eif2ak3, Eif2c2, Eif4g1, Eli2, Elp2, Elp3, Eny2, Ercc3, Ercc6, Etv3, Ezh1, F2r, Fam58b, Fntb, Foxj3, Gabpa, Gclc, Gm10094, Gtf2e1, Gtf2e2, Gtf2i, Hdac2, Hexb, Hivep1, Hmga1-rs1, Hnrnpa2b1, Hnrnpab, Hsf2, Huwe1, Ilf3, Ino80c, Insig1, Insr, Jazf1, Jmjd1c, Kcnip3, Kdm3a, Kdm5d, Khdrbs3, Lanc2, Ldb2, Mbd1, Mbd5, Meaf6, Med17, Med4, Mef2a, Mlh3, Mlt11, Mta3, Mterfd1, Myc, Myef2, Ncbp1, Ndn, Nfx1, Ngly1, Npat, Pcbbd2, Pcbp2, Pex1, Phc3, Picalm, Pkia, Pnrc2, Polh, Polr2a, Polr2b, Polr3b, Prckb, Prkdc, Prmt2, Prnp, Prpf8, Pspc1, Pten, Rad21, Rad23a, Rbbp5, Rfx7, Rprdl1a, Scai, Setdb1, Sfmbt1, Smad2, Smarcd1, Snappc1, Snx6, Sox5, Sp3, Stat1, Supt7l, Suz12, Tada1, Taf1, Taf13, Taf7, Tbk1, Tbl1xr1, Tceal8, Tcf20, Tir3, Tmpo, Tmsb4x, Tnks, Top1mt, Top2b, Topbp1, Traf7, <b>Trim33</b> , Tsg101, Ubqln2, Ubr2, Usp47, Usp7, Wac, Wdr61, Wdr77, Xrcc2, Xrcc4, Zbtb1, Zbtb33, Zbtb41, Zbtb6, Zfml, Zfp101, Zfp169, Zfp189, Zfp191, Zfp192, Zfp280d, Zfp317, Zfp322a, Zfp382, Zfp386, Zfp397, Zfp418, Zfp445, Zfp507, Zfp51, Zfp518a, Zfp518b, Zfp53, Zfp58, Zfp597, Zfp60, Zfp605, Zfp654, Zfp68, Zfp719, Zfp763, Zfp770, Zfp780b, Zfp791, Zfp82, Zfp84, Zfp871, Zfp874a, Zfp874b, Zfp948, Zfp949, Zfp958, Zhx1, Znym2
<b>Cluster 11</b>	Abca2, Actl6a, Bcl10, Bmp5, Cat, Ccna2, Chd1l, Creb3l2, Ctnnb1, Dynl1, Etv5, Fbxo21, Foxj1, H3f3a, Id4, Il33, Irak4, Kat2b, Map3k2, Mcm2, Mcm4, Mcm6, Med10, Mkl2, Nab1, Nck2, Nedd4, <b>Nfib</b> , Pcnal-ps2, Prickle1, Rad51, Ramp3, Rbmxi1, Rnasel, Runx1t1l, Rxra, Rybp, Sall1, Sik1, Sirt2, Smad7, Tfdp1, Trib1, Trp53inp2, Whsc1, Xrcc3, Zfhx3, Zfp266, Zfp551, Zmiz1
<b>Cluster 12</b>	Bcl11b, Bmp3, Cdon, Crym, Erbb2, Fgf10, Fgfr2, Foxo6, Gabpb2, Gm98, Id3, Itgb3bp, Jup, Klf4, Klf10, Lass2, Lbr, Litaf, Med12, Mif4gd, Nfe2l3, Olig1, Olig2, Otx1, Pbx1, Phox2a, Pou6f2, Prcko, Prox1, Rcbbt1, Rhog, Rps6ka5, Rsc1a1, Setdb2, Skil, Sox10, Sox8, Stat6, <b>Tbr1</b> , Tie4, Traf5, Trf, Xpo1, Zfpm2, Zkscan16

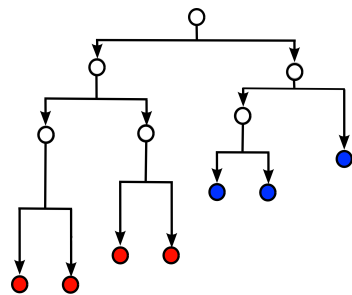
**Table S1. Transcription factor clusters** 751 Transcription factors (Belgard et al., 2011) were clustered according to the distribution of their normalized expression values across layers 6a, 6b, 5, 4 and 2-3 . The transcription factors of each cluster that were chosen as representative examples for Figure 5 are highlighted in bold.

## 1113 **10. Supporting Information: Figures**

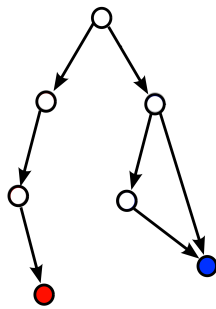


Figure S1

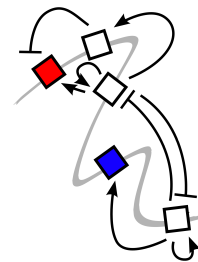
**A**



**B**

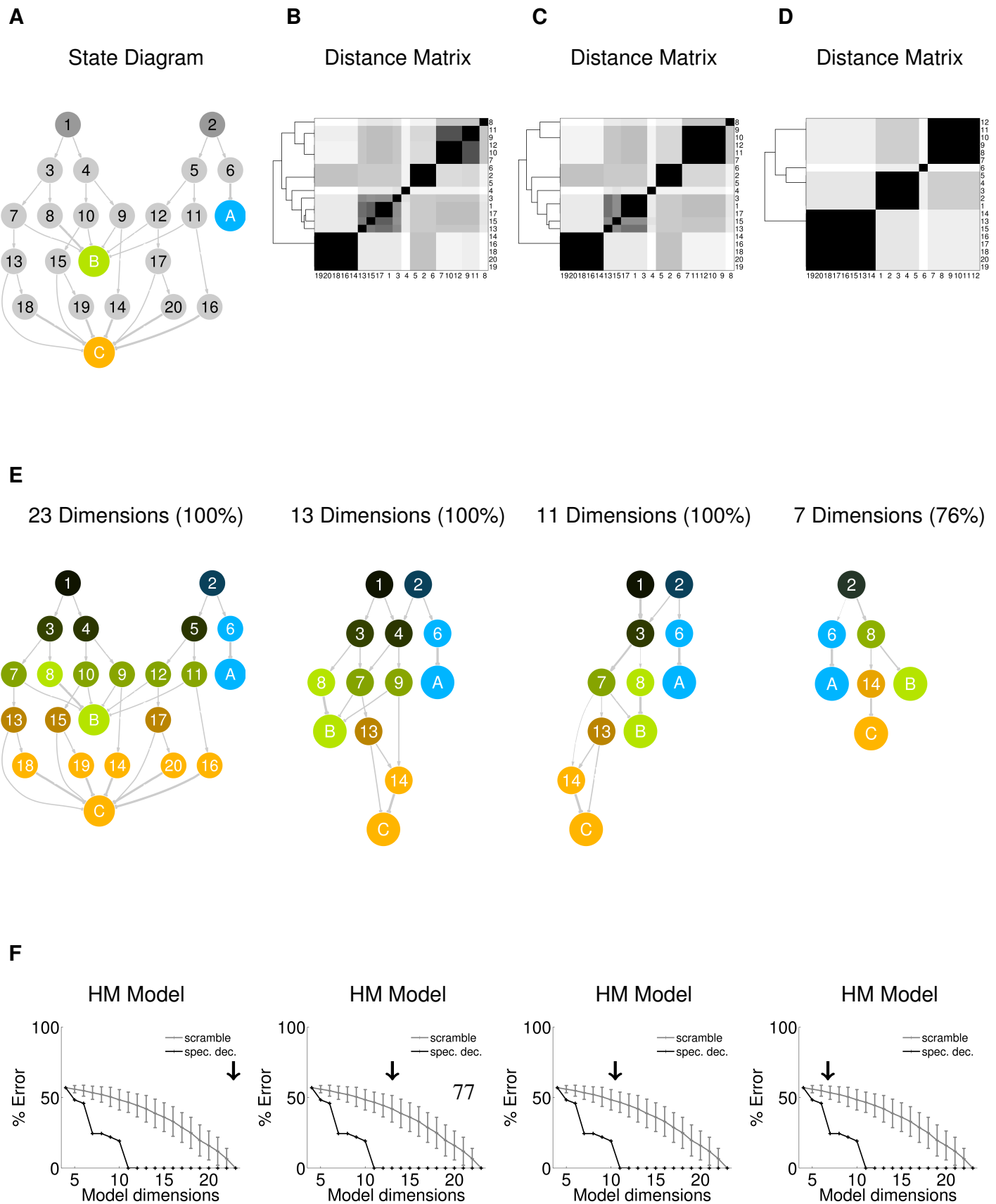


**C**



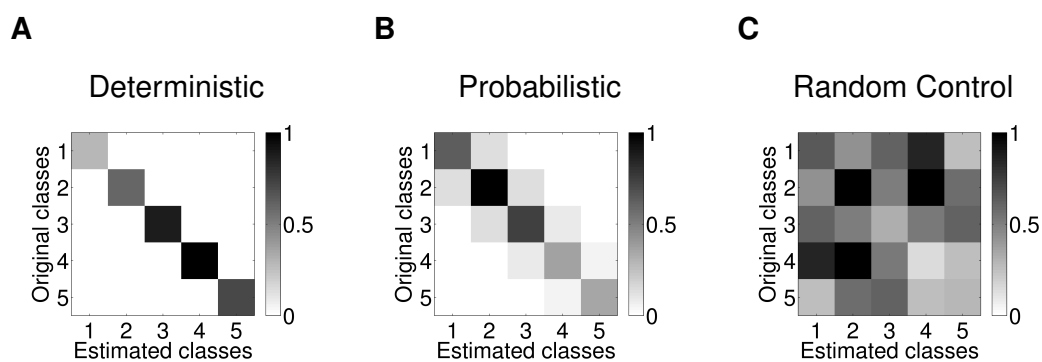
**Figure S1. Aspects of biological development.** The process of development can be understood in terms of three complementary models **(A)** The cell lineage tree describes the mitotic process rooted in a given precursor. Each cell divides symmetrically or asymmetrically to produce two similar or dissimilar daughter cells. Colors denote the different fates of terminal cells. **(B)** A phenotypic model of the possible states taken by cells of lineage tree. Each node represents a cell state that is characterized by a vector of observable features. Each edge represents a possible transition route between states. Colors denote the features expressed by terminal cell. **(C)** A genotypic model that is the mechanism underlying the lineage tree description, or the state diagram description. Each cell state is encoded by the expression of a subset of genes (squares) layed out on the DNA (gray line). The progression through the successive cell states of the lineage tree is controlled by gene interactions (black lines), and the degree of asymmetrical of cell division and gene interactions (black lines). These interactions may be positive (arrow) or negative (plate) with respect to their target genes. Colors represents genes linked with a particular terminal cell type.

**Figure S2**



**Figure S2. Reduction of State Diagram to lower dimensionality.** (A) State diagram of example lineages (as Figure 2B). Nodes represent cell states, arrows state transition probabilities. States are labeled according to 3 observed features:  $A = \langle 1,0,0 \rangle$  (blue),  $B = \langle 0,1,0 \rangle$  (green),  $C = \langle 0,0,1 \rangle$  (orange), and  $\# = \langle ?,?,? \rangle$  (gray) for states with hidden features. Initial states are depicted in dark gray. (B-D) State clustergrams of computed distance between every state pair with dimensions  $D = 23$ ,  $D = 13$ ,  $D = 11$ , and  $D = 7$  (percentage of data represented in parenthesis). Dendrograms indicate hierarchical binary linkage of states. (E) Spectral label propagation on models, where each hidden node is colored according to its estimated feature distribution. (F) Model error as percentage of the correct final cell state distribution for spectral clustering (black) versus random model (gray, standard deviations for 100 trials). HM, Homogeneous Markov model.

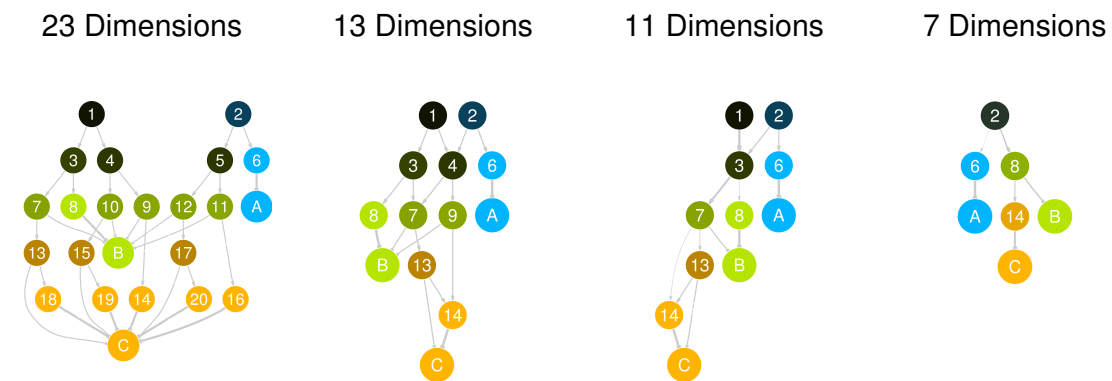
**Figure S3**



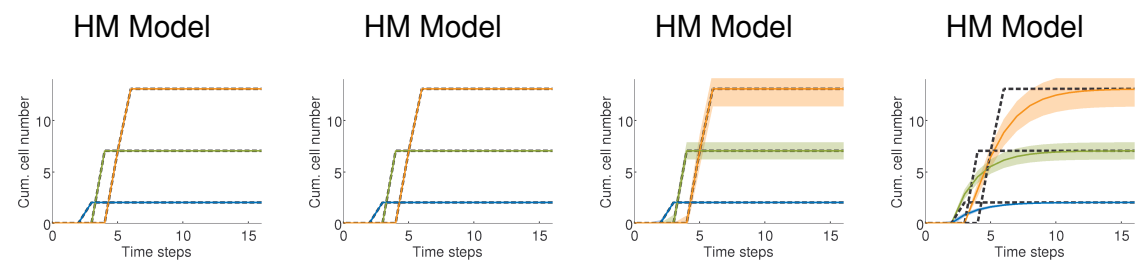
**Figure S3. Classification performance of spectral clustering.** The ability of spectral clustering to recover the correct Markov branching process was assessed on 100 lineages generated with 10 random 5-state models. Spectral clustering assigns a unique class to each cell, which is then compared to the known model class. **(A)** Confusion matrix of spectral clustering on deterministic model ( $0 \pm 0\%$  classification error). **(B)** Confusion matrix of spectral clustering on probabilistic model ( $20.3 \pm 17.8\%$  classification error). **(C)** Confusion matrix of random model ( $88.2 \pm 18.7\%$  classification error).

**Figure S4**

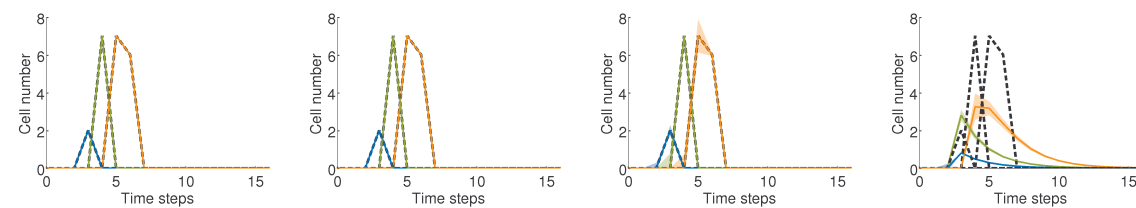
**A**



**B**



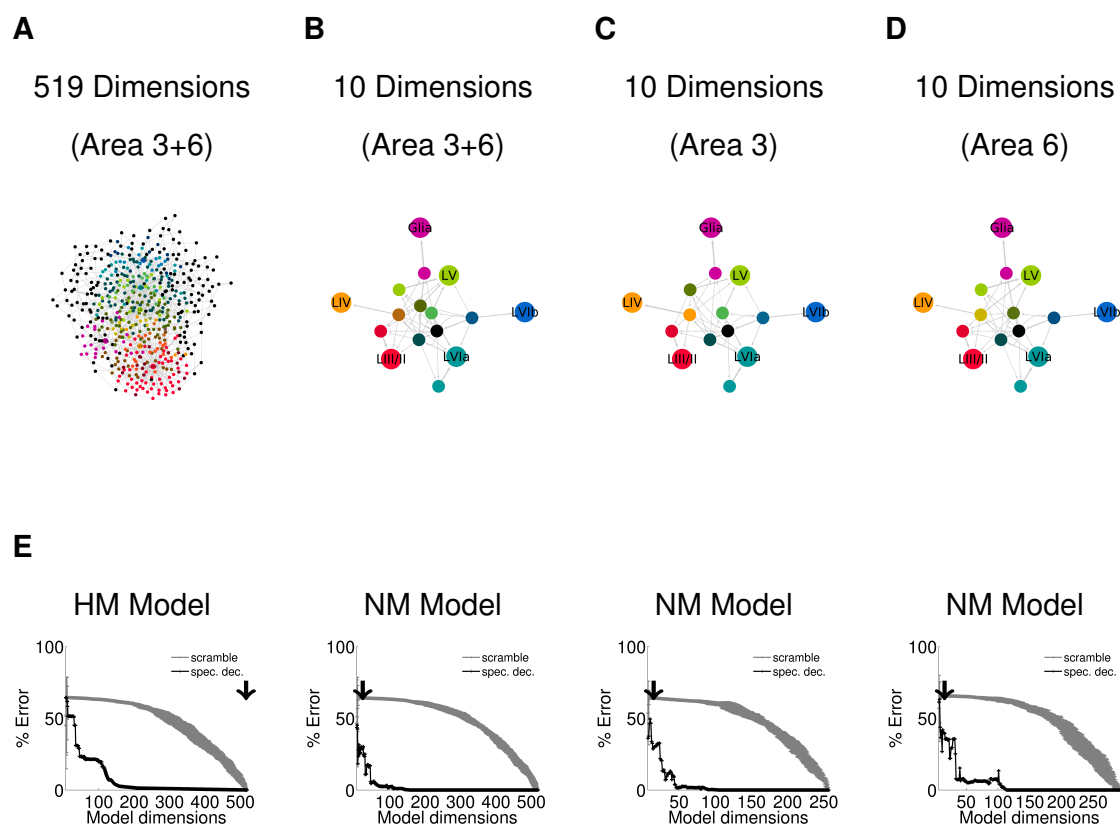
**C**



**Figure S4. Cell type distributions generated by a State Diagram of decreasing dimensionality.** (A) A State Diagram of an example sublineage is progressively reduced from dimension  $D = 23$  to  $D = 13$ ,  $D = 11$ , and  $D = 7$ . Nodes represent cell states, arrows state transition probabilities. (B) Output generated by Hidden Markov implementation of a State Diagram. Mean cumulative number of differentiated cells produced at each time step. (C) Mean instantaneous number of differentiated cells produced at each time step. Dashed lines, original distribution; colored lines, model distribution; shaded area, standard deviation. The  $D = 7$  model fails to capture the original data.

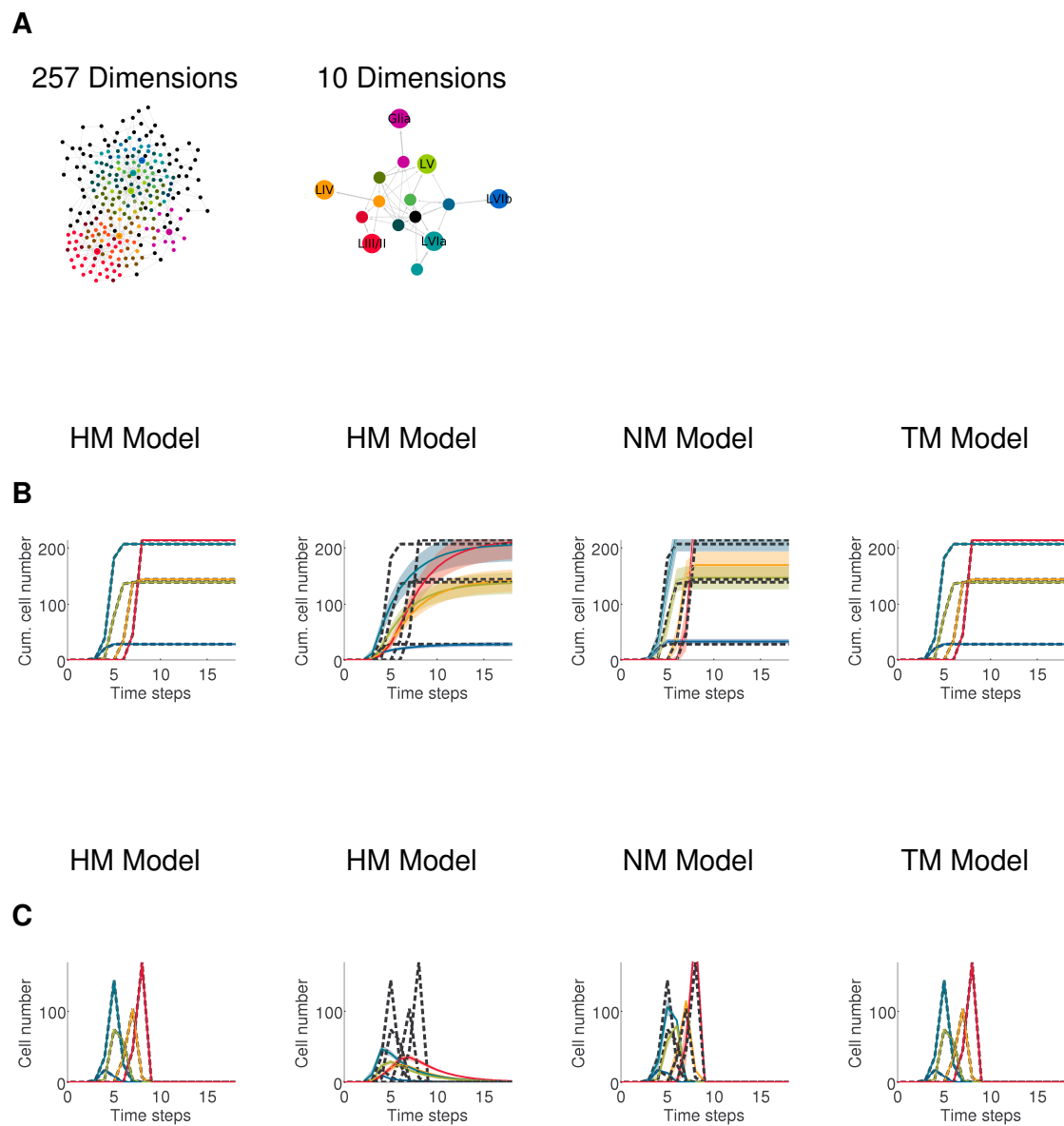


**Figure S5**



**Figure S5. State Diagrams areas 3 and 6 combined, and separated.** (A) 519-dimensional State Diagram of combined lineages for area 3 and 6. Nodes represent cell states, arrows state transition probabilities. (B) Combined SD reduced from  $D = 519$  to  $D = 10$  (area 3 and 6). (C)  $D = 10$  SD for area 3 alone. (D)  $D = 10$  SD for area 6 alone. Cell states: Layer 6b, blue; Layer 6a, sea green; Layer 5, green; Layer 4, orange; Layer 2/3, red; Glia, pink; and Unknown, gray. (E) Performance (% error against original data) of stochastic generator models (black traces) corresponding to the SDs above. The performance of the stochastic models is compared against a model free random control (grey traces). HM, Homogeneous Markov model; NM, Non-Homogenous Markov Model. Model dimension indicated by black arrow.

**Figure S6**



**Figure S6. State Diagrams and model generated cell distributions for cortical area 3. (A)** Original State Diagram  $D = 257$  and its reduced  $D = 10$  version for cell lineages in cortical area 3. Nodes represent cell states, arrows state transition probabilities. Cell state colors are the same as for Figure S5. **(B)** Generation of cells by various stochastic models. Mean cumulative number of differentiated cells produced at each time step. **(C)** Mean instantaneous number of differentiated cells produced at each time step. Dashed lines, original distribution; colored lines, model distribution; shaded area, standard deviation. HM, Homogeneous Markov model; NM, Non-homogeneous Markov model; TM, Time-dependent Markov model. Low-dimensional HM model fails to capture the data, whereas TM performs well.

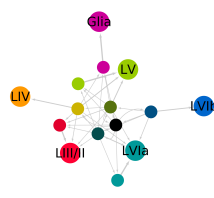
## Figure S7

**A**

292 Dimensions



10 Dimensions



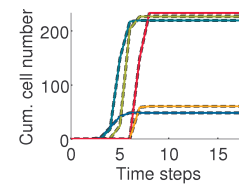
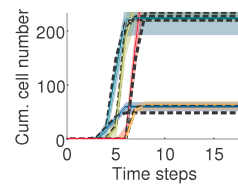
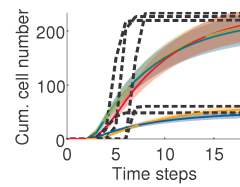
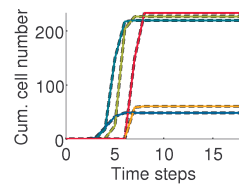
HM Model

HM Model

NM Model

TM Model

**B**



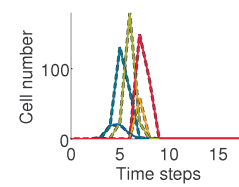
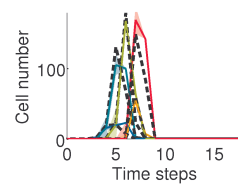
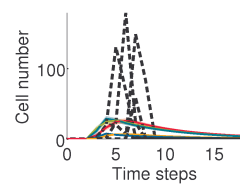
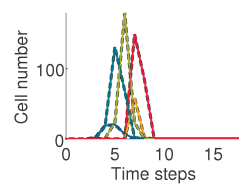
HM Model

HM Model

NM Model

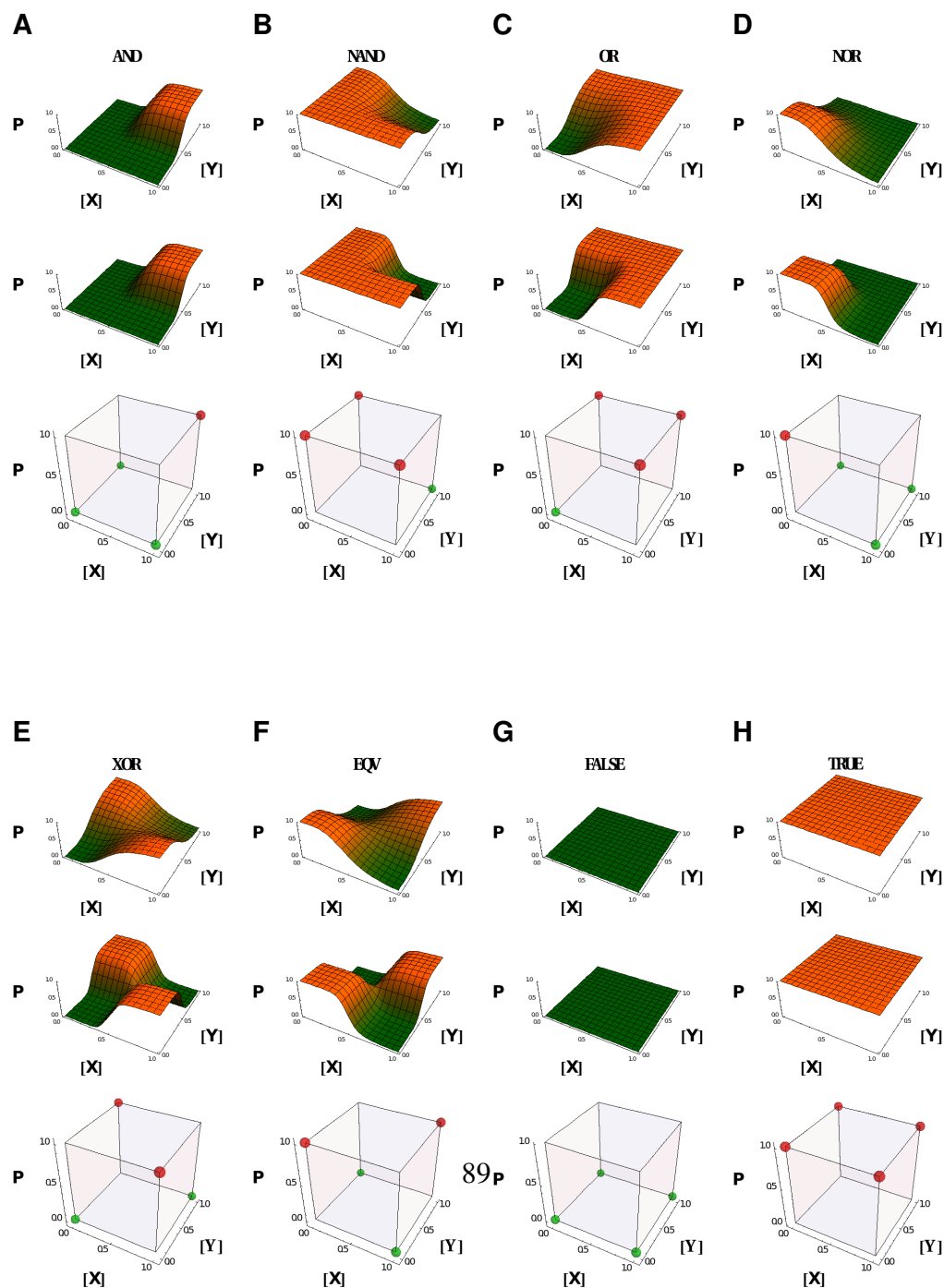
TM Model

**C**



**Figure S7. State Diagrams and model generated cell distributions for cortical area 6. (A)** Original State Diagram  $D = 292$  and its reduced  $D = 10$  version for cell lineages in cortical area 6. Nodes represent cell states, arrows state transition probabilities. Cell state colors are the same as for Figure S5. **(B)** Generation of cells by various stochastic models. Mean cumulative number of differentiated cells produced at each time step. **(C)** Mean instantaneous number of differentiated cells produced at each time step. Dashed lines, original distribution; colored lines, model distribution; shaded area, standard deviation. HM, Homogeneous Markov model; NM, Non-homogeneous Markov model; TM, Time-dependent Markov model. Low-dimensional HM model fails to capture the data, whereas TM performs well.

Figure S8

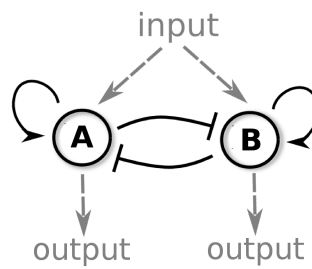


**Figure S8. Combinatorial transcription logic.** Cis-regulatory constructs can implement conventional canalizing logic gates **(A)** AND, **(B)** NAND, **(C)** OR, **(D)** NOR and non-canalizing **(E)** XOR, **(F)** EQV, **(G)** FALSE, **(H)** TRUE. The z-axis represents the output partition function  $P$  given  $[X]$  and  $[Y]$ . The computation depends on the steepness of the sigmoidal function  $H$ , ranging from (top to bottom row) continuous, approximately Boolean and discrete Boolean.

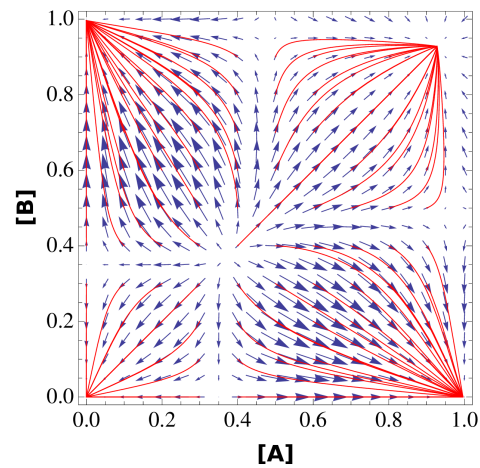


## Figure S9

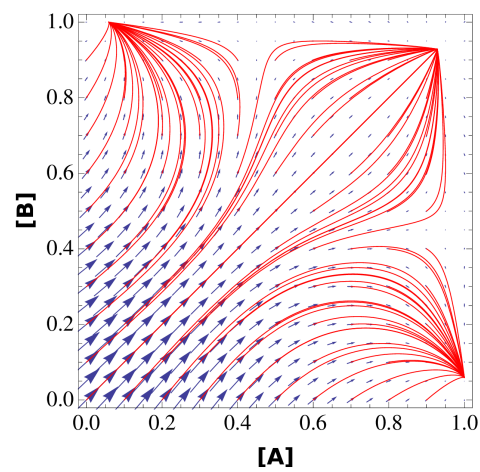
**A**



**B**



**C**



**Figure S9. Dynamics of a 2-dimensional genetic switch.** (A) Scheme of subnetwork with mutual inhibition between two transcription factors  $A$  and  $B$ , each with positive feedback; an external input  $I$ ; and two outputs. (B) Vector field representing the gradient direction as a function of concentrations  $A$  and  $B$ , for switch without input ( $I = 0$ ). The system has 4 attractor states, which means that the attractor states at high concentrations have hysteresis. (C) Vector field representing the gradient direction as a function of  $A$  and  $B$  for switch with input  $I = 1$ . Attractors at either high  $A$  or  $B$  represent downstream differentiation pathways. Red traces are simulated trajectories from various initial points.